SZEMERÉDI'S THEOREM: AN EXPLORATION OF IMPURITY, EXPLANATION, AND CONTENT

PATRICK J. RYAN

University of California, Berkeley

Abstract. In this paper I argue for an association between impurity and explanatory power in contemporary mathematics. This proposal is defended against the ancient and influential idea that *purity* and explanation go hand-in-hand (Aristotle, Bolzano) and recent suggestions that purity/impurity ascriptions and explanatory power are more or less distinct (Section 1). This is done by analyzing a central and deep result of additive number theory, Szemerédi's theorem, and various of its proofs (Section 2). In particular, I focus upon the radically impure (ergodic) proof due to Furstenberg (Section 3). Furstenberg's ergodic proof is striking because it utilizes intuitively foreign and infinitary resources to prove a finitary combinatorial result and does so in a perspicuous fashion. I claim that Furstenberg's proof is explanatory in light of its clear expression of a crucial structural result, which provides the "reason why" Szemerédi's theorem is true. This is, however, rather surprising: how can such intuitively different conceptual resources "get a grip on" the theorem to be proved? I account for this phenomenon by articulating a new construal of the content of a mathematical statement, which I call *structural content* (Section 4). I argue that the availability of structural content saves intuitive epistemic distinctions made in mathematical practice and simultaneously explicates the intervention of surprising and explanatorily rich conceptual resources. Structural content also disarms general arguments for thinking that impurity and explanatory power might come apart. Finally, I sketch a proposal that, once structural content is in hand, impure resources lead to explanatory proofs via suitably understood varieties of simplification and unification (Section 5).

§1. Introduction.

1.1. Our question and two crucial distinctions. I would like to explore a very general philosophical question restricted to a context in which it becomes particularly pressing. Our question is, "What kinds of conceptual resources are best suited for producing an explanation of a given fact?," and our context is pure mathematics. The reason that this question is especially important for mathematics is elegantly summarized by the distinguished combinatorialist, Gian-Carlo Rota:[1]

---

[1] Let me address at the outset an important methodological worry raised by an anonymous referee concerning the value of mathematicians' judgements about explanatory power, fruitfulness, etc. for the philosophy of mathematics. Concisely put, the worry seems to be that such judgements, being the informed opinions of a few practitioners, have little to tell philosophers attempting to articulate more objective features of mathematics itself.

> The mystery as well as the glory of mathematics lies not so much in the fact that abstract theories turn out to be useful in solving problems but in the fact that—wonder of wonders—a theory meant for one type of problem is often the only way of solving problems of entirely different kinds, problems for which the theory was not intended. These coincidences occur so frequently that they must belong to the essence of mathematics. No philosophy of mathematics shall be excused from explaining such occurrences ([55], p. 114).

The history of mathematics bears out Rota's "conceptual convergence" claim in striking fashion.[2] One of its most famous and elementary instances is Descartes' invention of analytic geometry whereby problems in Euclidean geometry were solved by representing them via algebraic equations, solving the equations, and translating the algebraic solutions back into geometrical terms.

When one examines the contemporary mathematical landscape, does Rota's claim hold? The answer is a resounding yes. As he says, the "coincidences" are so frequent that it seems they are not coincidental! The inter-penetration of intuitively different areas of mathematics, like the famous Cartesian example, is now a fact of life. This occurs at both a local and global level. Locally, we often witness techniques intended for the solution of one problem used in the solution of another entirely different one. Globally, there are now many explicitly "hybrid" *subfields* of mathematics, e.g., analytic number theory, algebraic topology, algebraic geometry, functional analysis, etc.[3]

One way to describe these mathematical developments is via a distinction between "impure" and "pure" mathematics.[4] Mathematicians commonly describe a proof of a theorem as "pure" if it uses only what is "intrinsic" or "close" to the theorem. On the other hand, a proof is impure if it draws on what is "extrinsic," "distant," or "foreign"

To my mind, this objection is only problematic for the enterprise of, say, developing an account of mathematical explanation, if the individual judgements of mathematicians are taken as dogma. It should be clear that, whenever I cite a mathematician's evaluation of a result, etc. in this paper, I take this evaluation as *(potentially defeasible) evidence* for a philosophical claim being made. Indeed, it seems to me a reasonable methodological principle to utilize the judgements of those who know mathematics most profoundly as starting points for philosophical theorizing. These judgements can then serve as a bridge to more robust and objective accounts of mathematical phenomena. (One might here think of Aristotle's practice of collecting *endoxa*.) Finally, it bears keeping in mind that embracing a radical skepticism concerning the value of such judgements serves to block philosophical investigation into areas that might yield interesting results. Indeed, consider how much worse off the philosophy of science would be if all evaluations concerning explanatory power, fruitfulness, unification, etc. in physics, biology, and chemistry were ignored. These methodological remarks are quite general: what counts as an adequate starting point for the construction of any philosophical theory? There is no obvious answer to this question, and thus I believe we should use—and carefully weigh— everything at our disposal.

2 While Rota addresses theories that provide the *only* means for solving some particular problem, I consider weaker cases in which particular theories or concepts "meant" for problems of type X provide *alternative* and epistemically advantageous means of solving problems of intuitively different type Y.

3 In this paper, I am primarily concerned with the "local" case where intuitively foreign techniques are used to prove a theorem of ostensibly different conceptual type (impure proofs). However, I also take the existence of many hybrid mathematical subfields to be evidence for the general success of locally impure techniques.

4 For a more detailed treatment, see: [3–5, 15, 16].

to the theorem. Thus, proving a theorem of plane geometry using Descartes' analytic techniques produces an impure proof, while doing so via planar methods of Euclid's *Elements* produces a pure proof. Similarly, proving a number-theoretic statement using complex functions and their properties produces an impure proof (as in classical analytic number theory), while doing so with merely arithmetical properties produces a pure proof. Though these are but a few examples, a cursory glance at contemporary mathematical practice indicates that impure techniques and proofs abound.

The pure-impure distinction[5] can (and should) be precisified further, which amounts to making more precise the distance measure[6] implicit in purity ascriptions. One construal of this measure is that of "elemental closeness": a proof draws only on what is simpler or more elementary[7] than the theorem. Another is that of "topical closeness": a proof draws only on what belongs to the content of the theorem or what the theorem is about. Each metric then induces a purity constraint. This coarse distinction will suffice for now, but we shall see that things are much more complex in practice.

Ascriptions of purity and impurity provide us with a useful way of carving up the conceptual content of mathematical techniques and proofs. Thus, exploiting this distinction, our central question can now be stated as, "Are pure or impure techniques better suited for producing an explanation of a mathematical fact?" But now, one naturally asks, what is an explanation of a mathematical fact? To make our investigation tractable, a few more restrictions are in order. First, I consider explanations internal to mathematics, i.e., where one mathematical fact is used to explain another.[8] Second, I consider a "local" conception of mathematical explanation insofar as explanatory power is construed as a property of proofs.[9] The question of what constitutes an explanatory proof is, in and of itself, an expansive one that has begun to attract more attention in recent years.[10] Nonetheless, I shall neglect these discussions for now and select an ancient starting point that gets at a second important distinction.

In *Posterior Analytics*, A.13, Aristotle distinguishes between demonstrations "of the fact" (ὅτι; lit. "the that") and demonstrations "of the reasoned fact" (διότι; lit. "the why" or "because of which"). That is, in mathematics we have proofs that show *that* a theorem is true and proofs that show *why* a theorem is true.[11] This is, more or less, the distinction between non-explanatory, though perfectly cogent, proofs and explanatory

---

5 My statement of this distinction and its precisification follows [3].

6 Please note that whenever phrases like "distance measure" or "metric" are used to describe the "closeness" of concepts involved in purity ascriptions, these phrases are to be understood as useful shorthand. I am not committed to the claim that anything so precise as a metric can be assigned to such conceptual comparisons.

7 Of course, it is then incumbent on one to make (more) precise what is meant by "elementary." *Mutatis mutandis* for topical purity and content.

8 Thus distinguishing my discussion from "external" mathematical explanations of physical phenomena.

9 This serves to distinguish my discussion from a "global" conception of explanation in which explanatory power is a property of theories or frameworks. Similarly, I am mostly concerned with impurity as a property of proofs. See [46] for a discussion of both the external-internal and local-global distinctions in the context of mathematical explanation.

10 For a very nice survey on the topic of mathematical explanation see [44] and, again, [46]. See also the recent book by Lange [41].

11 Aristotle considers not only mathematics but axiomatized sciences in general. I say "axiomatized" because, even though Aristotle employs examples from Greek sciences that were not axiomatized, e.g., astronomy, the theory of explanation provided in the

proofs. Such a distinction[12] helps to make good sense of mathematics as it is actually practiced, especially the fact that theorems, from the most elementary to the most complex, are often proved multiple times and in various ways.[13] Explanatory proofs are supposed to generate some understanding of the result being proved, while non-explanatory proofs merely give one warrant for the truth of the theorem; explanatory proofs provide "the why," while non-explanatory proofs merely provide "the that." However, I believe more than this is required for a proof to be explanatory. That is, an explanatory proof must also indicate *how* the "reason why" manifests itself and tell some intelligible story as to how it interacts with other elements of the proof. Thus, in order to classify a proof as explanatory, it is not quite sufficient to answer the why-question latent in some theorem: one must also show how the answer to the "why" manifests itself.[14]

We now wish to understand how the distinctions between pure versus impure proofs and explanatory versus non-explanatory proofs map on to one another. In particular, as will emerge below, we wish to understand how it is possible that impure proofs could ever be explanatory. I emphasize once more that this question is of general interest and significance: how is it possible that concepts from some domain X can be used to answer a very specific why-question about concepts of intuitively different domain Y?[15] My claim is that, not only is it possible for this to happen in mathematics, it actually does happen and with great frequency. One might then say that *impurity is oftentimes essential for mathematical explanations*. This has not been a popular view; indeed, as I consider below, there are many eminent philosophers who have thought that *purity* and explanation go hand-in-hand (Field, Aristotle, and Bolzano). There are also hints in recent literature of the claim that there is no systemic relationship between impurity (or purity) and explanation.[16] I believe that both groups have significant difficulty in accounting for a crucial piece of data about contemporary mathematics: the pervasiveness of intuitively impure techniques and the proliferation of subfields of mathematics that are impure by design. Indeed, if, say, purity and explanation are naturally associated, and if explanatory proofs are recognized and valued by the mathematical community,[17] why have different areas of mathematics

---

Posterior Analytics proceeds from indemonstrable axioms and scientific first principles (*archai*; ἀρχαί).

[12] I employ the distinction independent of Aristotle's many conditions on when a syllogism counts as a demonstration, e.g., the premises of such a syllogism are immediate, better known than the conclusion, etc. See *An. Post.*, A.2, 71b20 for the statement of these conditions.

[13] Some examples that immediately spring to mind: the Pythagorean theorem, quadratic reciprocity, the Riemann–Roch theorem, the Prime Number theorem, and, of course, the theorem considered in this paper. For reflections concerning why it is that theorems are re-proved see [13].

[14] This talk will be made more precise once I introduce my case study below. See also the conclusion in which I briefly compare the impure (ergodic) and pure (combinatorial) proofs of Szemerédi's theorem.

[15] My discussion of Field below should elucidate this nicely. One might think that the use of mathematics in the physical sciences is an instance of two rather different conceptual domains interacting in an explanatory fashion.

[16] At times, this idea seems to be present in [41].

[17] One can find evidence for this claim in [43, 44]. See also my remark above in footnote 13 about the reproving of theorems.

not become atomized, at least more so than we now witness?[18] Why do we utilize seemingly "foreign" concepts so often? And if we think that there is no association whatsoever between impurity (or purity) and explanation, the same questions arise. Both opponents have great difficulty in accounting for the contours of contemporary mathematical practice, whereas my proposal can begin to help make sense of these.

Let me close this section with an eloquent articulation of the impurity-explanation association[19] by Solomon Feferman:

> Abstraction and generalization are constantly pursued [in mathematics] as the means to reach *really satisfactory explanations* which account for scattered individual results. In particular, *extensive developments in algebra and analysis seem necessary to give us real insight into the behavior of natural numbers*. Thus we are able to realize certain results, whose instances can be finitistically checked, only by a detour via objects (such as ideals, analytic functions) which are much more "abstract" than those with which we are finally concerned ([18], p. 3; emphases my own).

It would be valuable, then, for philosophers of mathematics to develop some account of how these detours through the abstract and the impure can lead to "really satisfactory explanations."[20] This will be the aim of the paper. However, before entering into the details of my positive proposal, let me consider the reasoning provided by some of the most eminent opponents of my view.

1.2. Motivating my opponents. Here I consider the views of those who believe that: either (i) pure proofs provide better explanations than impure proofs (Field) or, more strongly, (ii) purity is required for an explanatory proof (Aristotle and Bolzano). I also consider more general reasons for thinking that purity might at least be associated

---

[18] An anonymous referee correctly notes that this point concerning the lack of atomization in mathematics is quite complex. For instance, when we work "locally," viz., when we ask if the proof of a *particular* theorem is pure, the answer to this question will depend upon the statement of the theorem itself. And, if this theorem is from a "hybrid" subfield, e.g., analytic number theory, then a proof of this theorem using techniques from analytic number theory will not count as topically impure, even though the subfield itself depends essentially upon impure methods. There are likely a number of interesting points to be made here; however, I avoid this issue in my case study because my original theorem is squarely combinatorial. As such, its pure and impure proofs are easily distinguished.

[19] Though the language of purity/impurity is not used, Feferman's claim can be translated very naturally into these terms. I also believe that Rota's quote expresses something very similar.

[20] Please refer to footnote 1 for methodological remarks concerning the import of Feferman's claims. Indeed, I hope it becomes clear in the course of the paper that such a "detour" through the abstract and impure for explanatory gain is not merely an invention of Feferman's but is, rather, a phenomenon that occurs not infrequently in mathematics. This is part of the work done by my explication of the case study below. Further, I invite my reader to ask why it is that a theorem (Szemerédi's) with a perfectly cogent combinatorial proof was then re-proved using an extremely different mathematical framework (ergodic theory). This is a phenomenon that calls out for philosophical analysis. Obviously, I believe this analysis is best understood in terms of explanatory power and impurity. One need not appeal to anything Feferman says (or anything Tao says, or Rota says, etc.) in order to reach this conclusion, but I do believe their words lend some credence to it.

with explanatory power (drawing on work by Arana and Morrison). I conclude by addressing the idea that there may be no association whatsoever between purity (resp. impurity) and explanatory power.

Turning first to the weaker claim, we see that Hartry Field has expressed a preference for explanations closely related to pure proofs in mathematics. As is well known, Field is a staunch nominalist and has provided the most systematic attempt to eliminate mathematical entities from scientific explanations.[21] Interestingly, he believes that one should want to be an eliminativist on grounds independent of anti-platonist scruples:

> For *even on the assumption that mathematical entities exist,* there is a *prima facie* oddity in thinking that they enter crucially into explanations of what is going on in the non-platonic realm of matter. It seems to me that the most satisfying explanations are usually 'intrinsic' ones that don't invoke entities that are causally irrelevant to what is being explained. 'Extrinsic' explanations are acceptable [...] but it is natural to think that for any good extrinsic explanation there is an intrinsic explanation that underlies it. [...] I regard the acceptance of an extrinsic explanation as ultimate as at least somewhat odd ([20], pp. 18–19).

I hasten to note that Field is considering a rather different context: the use of mathematics in explanations of physical phenomena.[22] Nonetheless, one might well wonder if there is something odd about extrinsic explanations in general, i.e., independent of a causal construal of the extrinsic-intrinsic distinction. Should we think that there is always an intrinsic explanation in the offing, i.e., an explanation appealing only to concepts endemic or "close" to the *explanandum*? In particular, is it odd to think that we might have mathematical explanations of mathematical facts that make essential appeal to impure considerations? Field does not offer any arguments for this general preference, and, obviously, I wish to resist such a move. Nonetheless, it is worth thinking about whether there is an intuitive strangeness to extrinsic (impure) explanations.

Turning now to the stronger claim, we find two very important proponents of the idea that explanatory proofs *require* purity. Let us start with Aristotle. A consequence of his theory of demonstration in the *Posterior Analytics*, taken to be his "official" theory of explanation,[23] is that

> [I]t is not possible to prove/demonstrate by crossing from another genus/kind, e.g., [to prove] something geometrical by means of arithmetic (*An. Post.* 75a38-9).[24]

This purity constraint follows from Aristotle's view that different sciences are distinguished by their subject matter, i.e., "the underlying kind" or genus of the science, and the fact that demonstrations "make clear" the *per se* properties and relations of

---

[21] See [19].

[22] So, really, I am articulating a quasi-Fieldean view. I do not wish to saddle Field with a view he does not in fact hold.

[23] See [12] for an excellent historical and philosophical discussion of the *Posterior Analytics* in which questions about the nature of explanation are taken up. See also the commentary in [9].

[24] Translation my own from the Greek text of [54].

each kind in question.[25] Thus, if we wish to provide a demonstration of some fact, i.e., an explanatory deductive argument with the given fact as conclusion, we cannot appeal to resources external to the genus in question.[26]

This might seem an unpalatable and overly strong consequence. Need this purity constraint then concern us, especially given its dependence upon much of the Aristotelian philosophical apparatus? Perhaps. In order to see why, we must briefly consider the mathematical and philosophical work of Bernard Bolzano. In 1817, Bolzano offered a spectacular vindication of the Aristotelian purity constraint on distinctly explanatory grounds. His paper, *Rein analytischer Beweis*, provided a "purely analytic" proof of the Intermediate Value Theorem[27] (IVT), i.e., without appeal to geometrical, and thereby impure, considerations.[28] I quote him at length since he gives a very clear expression of his desired purity-explanation relation:[29]

> The most common kind of proof [of this fact, i.e., the IVT] depends on a truth borrowed from geometry [...] There is certainly no question concerning the *correctness*, nor indeed the *obviousness*, of this geometrical proposition. But it is clear that it is an intolerable offense against *correct method* to derive truths of *pure* (or general) mathematics (i.e., arithmetic, algebra, analysis) from considerations which belong to a merely *applied* (or special) part, namely, *geometry*. [...] [I]f one considers that the proofs of the science should not merely be *confirmations* (*Gewissmachungen*), but rather justifications (*Begründungen*), i.e., presentations of the objective reason for the truth concerned, then it is self-evident that the strictly scientific proof, or the objective reason, of a truth which holds equally for *all* quantities, whether in space or not, cannot possibly lie in a truth which holds merely for quantities which are in *space* ([56], p. 160).

That is, if we wish to present the objective order of mathematical truths, if we wish to provide an explanatory proof of a given fact, we cannot, like Aristotle, cross genera.[30] The analytic proof of the IVT served as one step in Bolzano's larger project of systematizing analysis and geometry using only "explanatory" proofs. And it was not the end of the story: the subsequent rigorous development of mathematical analysis (independent of geometrical proofs) added even greater weight to the hypothesis that

---

[25] Here I am following the traditional reading of 75a38-b2. For a dissenting interpretation, see the recent article by Steinkrüger [58] in which the main lines of the traditional reading are also discussed.

[26] Aristotle does allow for "exceptions" to the purity constraint, e.g., one can demonstrate an optical fact using geometry.

[27] This theorem states: *A continuous function f defined on an interval $[a, b]$ attains all intermediate values, i.e., if $f(a) = \alpha$, $f(b) = \beta$, and $\alpha \leq \gamma \leq \beta$ is given, then there is some $c \in [a, b]$ such that $f(c) = \gamma$. Similarly, for $\beta \leq \gamma \leq \alpha$.* Technically, Bolzano proves a slightly more general result and derives the classical IVT as a corollary.

[28] See [50] for the original text of Bolzano's essay in German. See [56] for an English translation. See [10, 36, 42] for arguments that Bolzano puts forth a theory of *mathematical explanation*.

[29] We have to be rather careful when considering Bolzano's claims about purity and explanation. While Bolzano believes that every explanatory proof must be pure, he does not think that every pure proof is explanatory. See [42] for an extended discussion.

[30] Of course, Bolzano specifies the genera differently, e.g., "pure" vs. "applied" mathematics.

there is a close relationship between purity and explanation.[31] It appeared as though this development got at "the essence" of analysis, and this was only possible via appeal to pure techniques. Thus, the Aristotelian purity-explanation thesis should be taken seriously.

Futhermore, independent of particular philosophical programmes, there are intuitive reasons to believe that a purity constraint yields distinctly explanatory dividends.[32] First, a topically pure proof might give one better warrant to believe that the intended statement has been proved, rather than some different, albeit closely related, one (see [3], p. 208). Thus, if one thinks of an explanation as an answer to a why-question, an appeal to topical purity would facilitate an explanatory proof by ensuring that one is answering the intended question. Impure techniques, on the other hand, might subtly modify the question one is trying to answer. Second, one might think that, even if we keep the appropriate question in view, the use of impure resources will cause us to lose data essential for answering the question.[33] Namely, utilizing concepts from a different mathematical setting (especially a more "abstract" or infinitary one) might result in the loss of data germane for producing an explanatory proof of the original result.

Finally, I would like to briefly note that in the contemporary literature on both purity and explanation there is little mention of *any* systemic association between purity (respectively, impurity) and explanation.[34] I find this state of affairs rather odd. On the one hand, we might expect more contemporary supporters of the purity-explanatory power association in light of the historical arguments sketched above. On the other hand, given what we witness in contemporary mathematics, viz., the consistent use of impure techniques for some sort of epistemic gain, we might expect more support for an association between impurity and explanatory power. However, there is little mention of either.[35] In any case, given the data from mathematical practice, it is worth investigating the relationship between impurity and explanation in greater detail.

1.3. Schematic of the argument. I will attempt to vindicate and explicate the association between impurity and explanation via an analysis of Szemerédi's theorem on arithmetic progressions and various of its proofs. We shall see that radically impure resources (both elementally and topically impure) are used to provide a perspicuous

---

[31] Indeed, for Aristotle and Bolzano, purity is *necessary* for producing an explanatory proof (omitting the so-called "exceptions" that Aristotle notes). In what follows, I deal merely with associations between epistemic virtues, not necessary connections.

[32] I do not wish to deny that, independent of explanation, a purity constraint might yield other epistemic dividends. For instance, an elementally pure proof might make very efficient use of the information available. See [3] for further discussion.

[33] A very similar idea is expressed by Margaret Morrison in [48]. Here she convincingly argues that unification and explanation often come apart in scientific theories precisely because the mathematical apparatus responsible for the unification abstracts from causal data essential for explaining physical facts.

[34] For instance, [41] considers a number of examples of impure and explanatory proofs. One gets the sense, however, that Lange does not believe there is any deeper association between impurity and explanatory power. See, in particular, his remarks at ([41], p. 292).

[35] One exception is the discussion in [43] of Pringsheim's development of complex function theory. Pringsheim emphasized that an "elementary" development of the theory would yield understanding of the essence of the mathematics in question. This talk might be construed as hypothesizing a relationship between some conception of purity and explanation.

and explanatory proof of this theorem. That is, I provide a case study in which impure resources show both *why* Szemerédi's theorem is true and *how* the reason why materializes; thus, my example functions as a direct and important[36] counter-example to the purity-explanation association. I shall show how this is possible in light of the explanatory potential of pure proofs discussed above. Finally, as I have tried to stress, I believe the association of impure methods and explanation is a relatively systemic feature of contemporary mathematics, and this has not been paid sufficient attention by recent philosophical discussions. Indeed, though I present only one instance of the impurity-explanation association here, I believe it points towards this stronger thesis given the centrality and depth of Szemerédi's theorem and its proofs.

The paper proceeds in the following way. First, I introduce our main theorem and provide criteria of selection that serve to bolster confidence in both the philosophical conclusions drawn and their scope. This is followed by a brief outline of the impure proof in question. I continue with an explicitly philosophical discussion designed to answer the question, "How can impure techniques generate explanatory proofs?" I believe that an answer to this question turns upon the availability of a sufficiently fine-grained notion of the content of a mathematical proposition, which I call *structural content*. This discussion of content (Section 4) is the heart of the paper. Finally, I close by sketching a proposal that intuitively impure resources that share structural content with the theorem to be proved can provide an explanatory proof via simplicity and unification.[37]

§2. Our main theorem and criteria of selection.

2.1. A basic introduction to Szemerédi's theorem.
We consider an important and deep[38] result of additive number theory: Szemerédi's theorem (on arithmetic progressions). I begin with some basic definitions.

DEFINITION 2.1 (Arithmetic Progression). *For $k, r \geq 1$, a k-length arithmetic progression, written as $a, a + r, \ldots, a + (k-1)r$, is a sequence of k integers such that each element of the progression differs from its predecessor by precisely r. We call a the* base point *and r the* radius *of the arithmetic progression.*

EXAMPLE 2.2. *The set of numbers $\{1, \ldots, 9\}$ is an extremely trivial example of an arithmetic progression. We have $a = 1$, $r = 1$, and $k = 9$.*

An arithmetic progression of interest to computational mathematicians is one of the longest known arithmetic progressions of primes:[39]

---

[36] Regarding the importance of the example, see, in particular, my final criterion of selection in Section 2.3.

[37] Where these epistemic virtues must be precisified and suitably understood. In another paper, I discuss in greater detail what I mean by simplicity and unification and how they lead to explanation.

[38] See [2] for a recent philosophical discussion of Szemerédi's theorem. The main purpose of Arana's paper is to disambiguate various notions of "depth" in mathematical practice taking Szemerédi's theorem as example. I have learned much from Arana's perceptive discussion here (and elsewhere); however, I discuss rather different questions in this paper.

[39] See [29]. Note that the main result of this paper, i.e., the primes contain arbitrarily long arithmetic progressions, now called the *Green–Tao Theorem*, is non-constructive and thus merely proves the *existence* of arbitrarily long arithmetic progressions of primes.

EXAMPLE 2.3. *Take $a = 56211383760397$, $r = 44546738095860$ and $k = 0, 1, \ldots, 22$ for an arithmetic progression of primes given by $56211383760397 + 44546738095860k$.*

Note that oftentimes we deal with *arbitrarily long* arithmetic progressions, where this is commonly taken to mean finite arithmetic progressions of arbitrary length, i.e., without the specification of some k.

We can now state a finitary[40] version of our theorem:

THEOREM 2.4 (Szemerédi, finitary). *For every integer $k \geq 1$ and real number $0 < \delta \leq 1$, there is an integer $N(k, \delta) \geq 1$ such that for every $N \geq N(k, \delta)$, every set $A \subset \{1, \ldots, N\}$ of cardinality $|A| \geq \delta N$ contains at least one arithmetic progression of length k.*

This is equivalent, via a routine compactness argument, to an infinitary version, proved in 1975 by E. Szemerédi, thereby providing an answer to a long-standing conjecture of Turán and Erdős:[41]

THEOREM 2.5 (Szemerédi, infinitary). *Let A be a subset of the integers $\mathbb{Z}$ with positive upper density, i.e.,*

$$\delta(A) := \limsup_{N \to \infty} \frac{|A \cap [-N, N]|}{2N + 1} > 0. \qquad (2.1)$$

Then A contains arbitrarily long arithmetic progressions.

This theorem will be the main focus of our paper.[42] It tells us that, for some very large (positive density) set $A \subset \mathbb{Z}$, one will always find arbitrarily long arithmetic progressions in A. This is quite remarkable for a few reasons.[43] First, no assumptions are laid down on the set under consideration aside from its size. Thus, even if two sets A and B look entirely different (but are both sufficiently large), each will have arbitrarily long arithmetic progressions. For instance, A could be randomly generated and B could *itself* be an arithmetic progression (an infinitely long one), and yet both will have long arithmetic progressions contained in them. In fact, it is quite easy to extract arithmetic progressions from either of these extreme cases, i.e., very random and very

---

40 Intuitively, this can be taken to mean "involving only finite sets." See below for a more precise statement of the finitary versus infinitary distinction.

41 Both Szemerédi's theorem and the Green–Tao theorem on primes are special cases of the most general (and still open) Erdős–Turan conjecture for arithmetic progressions: Suppose that $A = \{a_1 < a_2 < \cdots\}$ is an infinite sequence of integers such that $\sum 1/a_i = \infty$. Then A contains arbitrarily long arithmetic progressions.

42 Szemerédi's theorem has a rich history, which I suppress here in order to make the paper more manageable. Let it suffice to say that it was first investigated as a conjecture that would imply the famous van der Waerden's theorem on arithmetic progressions. This latter theorem states that if the natural numbers are written as the disjoint union of finitely many sets, then there must be one such set that contains arbitrarily long arithmetic progressions. Van der Waerden's theorem, like Szemerédi's, has a pure combinatorial proof and an impure dynamical (topological) proof. See [25, 64, 65]. See [61] for a gentle introduction to some of the main moves in the ergodic proof of Szemerédi's theorem using the topological proof of van der Waerden's theorem as prolegomenon. I might have used the example of van der Waerden's theorem, but the epistemic advantages of impure and infinitary resources in that case are much less striking than in the more general case of Szemerédi's theorem.

43 See the very nice survey [28] for further discussion.

structured cases, though for entirely different reasons. Indeed, *every* (known) proof of Szemerédi's theorem relies on what Terry Tao has called a fundamental *dichotomy between structure and randomness*.[44] Very roughly speaking, this dichotomy can be exploited to exhaustively characterize any sufficiently large set, and thus prove that any such set has arithmetic progressions. The dichotomy will be an important aspect of our philosophical discussion as it has implications for the relationship between content, impurity, and explanation.

One might be unimpressed by the presence of arithmetic progressions in very large sets: *of course*, if you choose something large enough, you will find what you are after. However, one finds that other related patterns are quite easily avoided in large sets. The canonical example is:

EXAMPLE 2.6. *Consider the set of squarefree numbers, S, i.e., positive integers divisible by no other perfect square but* 1. *This is a sufficiently large set: its upper density is* $\delta(S) = 6/\pi^2$. *However, one can avoid geometric progressions* (*progressions of the form* a, ar, ar^2 *for* $r > 1$) *in S*.

And yet, Szemerédi's theorem says that one cannot avoid arithmetic progressions in large subsets of integers. Thus, the lack of *a priori* constraints on the sets considered and the ease of eliminating other patterns from large sets provide two reasons for thinking that Szemerédi's theorem is particularly "deep." Furthermore, it has an impressive relationship to further developments in mathematics; I will say more about this at the end of the section.

2.2. A radically impure proof of Szemerédi's theorem. Szemerédi's original proof was entirely combinatorial[45] and thus paradigmatically pure (see [59]). It is, however, incredibly intricate and has been judged, even by the best professional mathematicians working in this field, as "remarkably subtle" (see [61]). Indeed, it is quite rare, except in the most difficult and involved of proofs, to include a schematic diagram outlining the proof steps, but this is just what Szemerédi does (see Section 6).

In 1977, Hillel Furstenberg gave an incredibly different and highly impure proof of Szemerédi's theorem.[46] The first step was to show that Szemerédi's theorem (Theorem 2.5) is in fact equivalent to a theorem about the behavior of dynamical systems.[47] This dynamical analogue could then be proved using techniques from ergodic theory. (All of this will be made precise below.) The important thing to note at this point is that the impurity is both elemental and topical: elemental because Furstenberg's proof uses highly infinitary techniques to prove a result with explicit combinatorial, finitary content and topical because dynamical systems (more precisely, measure-preserving systems) are intuitively very different from mere sets of positive integers. I believe both aspects of the impurity contribute to the explanatory power of Furstenberg's proof.

Before turning to our criteria of selection, let me make the finitary–infinitary distinction more precise, as this will feature throughout the discussion. As mentioned

---

[44] See [60]. I discuss this further in the following section.
[45] Roughly, involving finite structures of natural numbers.
[46] See [22]. The clearest presentation of the proof was given a few years later in [24]. I primarily follow this latter paper.
[47] See Definition 3.1.

above, finitary mathematics can be understood to mean the "theory of finite sets." Indeed, we have the following result:[48]

THEOREM 2.7. *First-order Peano arithmetic* (PA) *is equivalent to* ZF⁻, *i.e., Zermelo-Fraenkel set theory without the Axiom of Infinity and with the negation of the Axiom of Infinity. Here "equivalent" means that* PA *and* ZF⁻ *are mutually interpretable*[49] (*and thus equiconsistent*).

On the other hand, one can understand infinitary mathematics to mean the "theory of infinite sets." This distinction can take on new contours in different contexts, e.g., finitary mathematics in Hilbert's Program is weaker than what I am calling finitary here. When appearing in the mathematical wild, rather than in more precise logical contexts, finitary mathematics employs notions like: the cardinality of finite sets (of course), upper and lower bounds of such sets, and measures of bounded sets. Infinitary mathematics then employs notions like: sequences, measurable sets and functions, general measurability and integrability, convergence, and compactness. This battery of concepts should suffice to make clear when and why a result is called either finitary or infinitary below.[50] To slightly belabor the point: Szemerédi's theorem (given in Theorem 2.4) is avowedly finitary, Furstenberg's proof of it is avowedly infinitary, and thus Furstenberg's proof is elementally impure.

2.3. Criteria of selection. Let me here provide some motivation as to why I believe Szemerédi's theorem is worthy of philosophical attention and why an analysis of it might yield impressive philosophical dividends.

First, Szemerédi's theorem does not *require* infinitary and impure techniques to prove it.[51] This sets it apart from now canonical number-theoretic examples that utilize resources far beyond what one would think necessary to prove them, e.g., Fermat's

---

[48] See the recent paper by Kaye and Wong [35] in which they investigate this result and locate some imprecision in the folklore. They indicate that the commonly cited relationship between PA and set theory is what I have given in Theorem 2.7; however, in order to show that the interpretations are inverses of one another (bi-interpretability) one must carefully axiomatize ZF. This is interesting but does not affect the substance of the distinction made here.

[49] The interpretation of PA in ZF⁻ is quite easy to see: interpret 0 as $\emptyset$ and the successor function as $x \mapsto x \cup \{x\}$. The converse is more surprising but can be done by coding finite sets with natural numbers. In particular, defining the membership relation $\in$ on the natural numbers as $n \in m$ iff the nth digit in the binary representation of m is 1 interprets ZF⁻. This interpretation is due to Ackermann in [1]. For instance, 55 serves as the code for the finite set $\{0, 1, 2, 4, 5\}$ since $2^5 + 2^4 + 2^2 + 2^1 + 2^0 = 55$.

[50] For a nice reflection on finitary vs. infinitary mathematics in analysis, from which I have drawn, see Chapter 1.3 of [62].

[51] Indeed, it is likely the case that once much proof-theoretic work has been done, *most* mathematics does not require the full strength of infinitary techniques. We should be careful not to confuse this descriptive claim with a normative one that mathematics *ought* to be this way. Though we can "get away with" proving many theorems from a relatively restricted mathematical universe, a great deal is lost when this is done. See [6, 18] for eloquent articulations of similar views. An interesting question to which I would like to provide a partial answer is: what exactly is lost when we perform proof-theoretic ontological and epistemological reductions? E.g., I am skeptical of Avigad's strategy in [7] to provide a purely constructive, finitary proof of the ergodic version of Szemerédi's theorem that is *as perspicuous* as the ergodic proof. Even if an explicit combinatorial version of Furstenberg's proof can be given, I posit that explanatory value will be lost.

Last Theorem, the Paris–Harrington theorem, and Friedman's finitary version of Kruskal's theorem. In attempting to elucidate the explanatory dividends of infinitary and impure methods in mathematics, it is crucial that I examine results that admit of multiple varieties of proof. This allows for a fruitful comparison of techniques, a comparison that (I hope) inclines one think that infinitary and impure methods, though not "required" *per se*, are very much an essential part of mathematical practice and explanation.[52]

Second, the ergodic proof of Szemerédi's theorem has been subjected to meta-mathematical analysis by Jeremy Avigad and Henry Towsner (see [7, 8, 63]). The metamathematical data available will allow us to make some of our observations more precise, e.g., the stage of the ergodic proof of Szemerédi's theorem where crucial explanatory work is done may involve axiomatically strong mathematics. Nonetheless, I do not believe that it is primarily the axiomatic strength of the ergodic results that drives the explanatory power of Furstenberg's proof. The axiomatic strength is, rather, a side effect of the move to "explanatorily appropriate" resources.[53]

Finally, Szemerédi's theorem is an absolutely central result in number theory, and its proofs, especially Furstenberg's, have generated impressive developments in diverse fields of mathematics.[54] I take these facts to be quite important because, to my mind, many philosophical analyses of mathematics are unsatisfying or fail outright because they extract conclusions from results that have no obvious mathematical significance.[55] This should then lead one to doubt the significance of the philosophical conclusions drawn. That is, if one is attempting to provide a philosophy *of mathematics*, the examples analyzed should be of acknowledged mathematical import.[56] Szemerédi's theorem, on the other hand, has been a locus of work in number theory and combinatorics for almost half a century, and, as we shall see, gave rise to entirely new sub-fields of mathematics. Finally, the conceptual approaches used to prove this theorem have shown astonishing applicability in the solution of other, long-standing questions, e.g., whether the prime numbers contain arbitrarily long arithmetic progressions.[57] Indeed, the full significance of these results is not yet understood.

---

[52] I fulfill this comparative promissory note more fully in a forthcoming paper. Here I consider only the ergodic proof in any detail.

[53] Section 4 is intended to provide a sense in which one mathematical domain may be "explanatorily appropriate" for proving results from another intuitively different domain.

[54] One should also note that there are other proofs of Szemerédi's theorem beyond the two I consider here. See the Fourier-analytic approach by Gowers in [26] and the hypergraph approach by Gowers, Nagle, Rődl, Schacht, and Skokan in [27, 49, 51–53].

[55] On the other hand, one should keep in mind that not every mathematically significant result is philosophically significant.

[56] This is not to say that simpler cases are of no use; indeed, in his [41], Marc Lange considers nothing beyond the reach of an undergraduate in mathematics and in so doing assembles an important collection of phenomenological data. However, at some point, philosophers have to engage with the "embarrassment of riches" offered by contemporary mathematics (see [44]). I think the philosophical gains will be similarly rich, and I hope that this case study inclines my reader to think this as well.

[57] As noted above, this was answered affirmatively by Ben Green and Terry Tao in [29]. It is very interesting that their entirely finitary argument is conceptually closer to techniques in infinitary ergodic theory than it is to techniques in quantitative analysis. Indeed, they rely upon methods akin to those employed in the ergodic proof of Szemerédi's theorem. This indicates, once more, the importance and depth of these methods.

§3. A summary of results in ergodic theory.

3.1. Furstenberg's ergodic proof of Szemerédi's theorem. Here I provide the conceptual contours of Furstenberg's ergodic proof of Szemerédi's theorem.[58] The summary here should contain all the mathematical material needed to engage with the discussion below.

First, a word on "ergodic theory." Broadly speaking, ergodic theory is the study of the long-term behavior of dynamical systems:

DEFINITION 3.1. *A* dynamical system *is a pair* (X, T) *where X is a set (or abstract space) and $T : X \to X$ is an invertible map (i.e., a map with well-defined inverse) operating on the elements of X. By $T^n x$ we understand T applied n times to some $x \in X$.*

Intuitively, we study the time-evolution of X under transformation T.

For example, we might consider the finite system (X, T) where X is a finite set and T a permutation of the elements of X. Alternatively, we might understand a dynamical system as the action of a group G on a set X.[59] Thus far, at this level of abstraction, little interesting can be said. However, applying a little more structure to dynamical systems will yield surprising results. *Classical* ergodic theory, which is what concerns us here, studies *measure-preserving systems*:

DEFINITION 3.2. *A* measure-preserving system *is a quadruple* $(X, \mathcal{B}, \mu, T)$ *where X is a set (or abstract space), $\mathcal{B}$ is a σ-algebra of subsets of X, $\mu : \mathcal{B} \to [0, 1]$ is a probability measure, namely, μ is additive and $\mu(X) = 1$, and $T : X \to X$ is a measurable transformation such that $\mu(T^{-1}(E)) = \mu(E)$ for all $E \in \mathcal{B}$, i.e., T is measure-preserving.*

These definitions, along with the information above on Szemerédi's theorem, should suffice for an outline of the ergodic proof.

Let us call the first stage in the ergodic proof of Szemerédi's theorem the *Correspondence Stage*. Furstenberg realized that Szemerédi's theorem would follow from a generalization of a famous recurrence result called *Poincaré Recurrence* in ergodic theory.[60] (In essence, a recurrence result describes how points in a dynamical system return close to themselves after sufficiently many iterations of map T.[61]) This generalization is now called *Furstenberg Multiple Recurrence* in his honor. Indeed, one can also show that Szemerédi's theorem implies Furstenberg Multiple Recurrence, thereby establishing a correspondence between the combinatorial and ergodic settings. Let me now state these results more precisely.

THEOREM 3.3 *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and let $E \in \mathcal{B}$ with $\mu(E) > 0$. Then for any $k \geq 1$,*

---

[58] I have primarily followed the survey [24]. See also Chapter 7 of the textbook by Einsiedler and Ward [17] as well as Tao's discussion in [61] and the expository paper [67]. I have benefited a great deal from all of these. Indeed, I claim no originality in the following technical presentation.

[59] Indeed, ergodic theory is naturally viewed as the study of measure-preserving group actions; however, we need not explicitly deal with algebraic techniques here. See Chapter 1 of [17] for a discussion of what kinds of results might count as ergodic.

[60] See [23] for a nice discussion of recurrence results.

[61] More precisely, we say that an element x of a dynamical system (X, T) is a recurrent point if, for some sequence $n_k \to \infty$, $T^{n_k} x \to x$.

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(E \cap T^n E \cap \cdots \cap T^{(k-1)n} E) > 0. \tag{3.1}$$

Then we can show

THEOREM 3.4 (Furstenberg Correspondence). *Furstenberg Multiple Recurrence and Szemerédi's theorem are equivalent.*

I will not prove these results here.[62] However, let me attempt to give my reader a sense as to how the various ideas fit together.

We begin with a problem about arithmetic progressions in the integers and wish to convert it to an ergodic setting. The basic idea here is that we can identify subsets $A \subset \mathbb{Z}$ with subsets of the ambient set or space X in the measure-preserving system. Similarly, we can identify functions on the integers with functions on the dynamical system. After this is done, the task of finding an arithmetic progression in some $A \subset \mathbb{Z}$ is in fact equivalent to finding an arithmetic progression in the subset of X identified with A. Let me try to make this a bit more precise. We can show, for our measure-preserving system $(X, \mathcal{B}, \mu, T)$, that there is a probability measure μ on X such that $\mu(E) = \mu(T^n E)$ for all $n \in \mathbb{Z}$ and measurable sets $E \subset X$ with E of positive measure. Then, the existence of some k-length arithmetic progression in $A \subset \mathbb{Z}$ (the assertion of Szemerédi's theorem) is equivalent to the existence of some $x \in X$ such that a recurrence pattern $x, T^n x, \ldots, T^{(k-1)n} x$ is in E.

In particular, we show that

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(E \cap T^n E \cap \cdots \cap T^{(k-1)n} E\right) > 0 \tag{3.2}$$

for $\mu(E) > 0$. This implies the recurrence claim above, i.e., $x, T^n x, \ldots, T^{(k-1)n} x$ is in E, since it shows that the intersection of sets $E \cap T^n E \cap \cdots \cap T^{(k-1)n} E$ is non-empty for infinitely many n. This concludes the Correspondence Stage of the proof.

The second stage of the proof involves showing 3.2 holds, i.e., proving Furstenberg Multiple Recurrence. I believe this is where the ergodic proof yields great explanatory dividends. In short, any proof of 3.2 requires decomposing the system under consideration into a structured component and a random component, which, by the Correspondence Stage of the proof, is equivalent to decomposing any $A \subset \mathbb{Z}$ into a structured set and a random set. Tao presents the situation with usual adeptness as a "fundamental dichotomy between structure and randomness." He continues,

> ...the reason for the truth of Szemerédi's theorem is very different in the cases when A is random, and when A is structured. These two cases can then be combined to handle the case when A is (or contains) a large (pseudo-)random subset of a structured set. Each of the proofs of Szemerédi's theorem[63] now hinge on a *structure theorem* which,

---

[62] Again, see [22], [24], and [17], Chapter 7.
[63] Of which, at the moment, (and not including generalizations) there are four: (i) combinatorial, (ii) ergodic, (iii) Fourier analytic, (iv) hypergraph. See footnote 54 for references.

very roughly speaking, asserts that *every* set of positive density is (or contains) a large pseudo-random[64] subset of a structured set ([60], p. 583).

I will have much more to say about structure theorems below. Returning to the proof sketch, it turns out that in the ergodic setting the requisite structure theorem and the dichotomy between structure and randomness are particularly explicit. Let's sketch how the structure theorem is applied in the ergodic setting.

The dichotomy presents itself as that between the periodicity of some $E \subset X$ under transformation T (structured case) or the mixing of E under transformation T (random case). If E is periodic, then this simply means $T^l E = E$ for some $l > 0$, i.e., we have a very obvious recurrence of E under T for some $l > 0$. Establishing 3.2 is then trivial as the summand will be $\mu(E)$ whenever n is a multiple of l. Even in the less well-behaved case of near-periodicity, we can show 3.2 in a very similar fashion. When all E for a space X are near-periodic, then we say the measure-preserving system is *compact*.

On the other hand, we get the random scenario when E has a "mixing" property. Intuitively, this can be thought of as follows: if E is some event in the probability space X, then T "mixes" the space randomly such that all events $\{E, T^1 E, T^2 E, ...\}$ are independent of one another. This yields 3.2 trivially as then each recurrence pattern is just $\mu(E)^k$. As in the structured case, even when we relax the mixing to the case of "weak-mixing," 3.2 can be proven. When all E in X have this weak-mixing property, we say that the system is a weak-mixing system. In either the compact or weak-mixing cases, we can prove our desired result quite easily.

However, not all systems are either compact or weak-mixing.[65] Nonetheless, provided that X is not completely weak-mixing (i.e., totally random), we can always find some structured component Y of X, which will itself give rise to a compact system. We can then analyze the map from $X \to Y$, called an *extension map*, which itself has structured or random behavior. If $X \to Y$ is a weak-mixing or compact extension,[66] then, even if X is neither of these, we can reduce the task of showing 3.2 holds in X to showing this holds in Y, and, since Y is compact, showing 3.2 holds in Y can be done easily. However, like systems, not all extensions are either weak-mixing or compact. If $X \to Y$ is some such intermediate case, then we can find some intermediate extension $X \to Y_1 \to Y$ such that $Y_1 \to Y$ is a compact extension. We can then pass to Y and show 3.2 holds. Thus, we must show that if 3.2 holds in Y_1 it holds in X. This process can be continued (transfinitely) by constructing a *tower* $X \to Y_\alpha \to \cdots \to Y_1 \to Y$

---

64 Just as with "random" and "structured," one must make explicit what is meant by "pseudo-random" in a particular context. Very generally, we can say that a pseudo-random set of integers resembles a random set of integers with similar density in terms of particular arithmetic statistics. For instance, in [29], Green and Tao construct a superset of the primes (the "almost primes"), which is pseudorandom in the sense that it satisfies a "linear forms condition" and a "correlation condition." These conditions correspond very closely to the ergodic notion of weak-mixing.

65 This differs from the topological proof of van der Waerden's theorem where all spaces under consideration will either be minimal (structured) or not. This fact is one reason why, as I noted above in footnote 42, the epistemic dividends in the simpler case of van der Waerden's theorem are much less impressive.

66 In which case we say that X is weak-mixing or compact relative to Y.

where each step $Y_{n+1} \to Y_n$ is a compact extension and $X \to Y_\alpha$ is weak-mixing. This is the essence of the Furstenberg Structure Theorem:

THEOREM 3.5 (Furstenberg–Zimmer Structure Theorem). *Let $(X, \mathcal{B}, \mu, T)$ be any measure preserving system. Then there is an ordinal α and a transfinite increasing sequence of factors*[67] *$(\mathcal{B}_\beta)_{\beta \leq \alpha}$ such that*:

1. *$\mathcal{B}_0$ is trivial (i.e., $\{\emptyset, X\}$).*
2. *For each successor ordinal $\beta + 1 < \alpha$, $(X, \mathcal{B}_{\beta+1}, \mu, T)$ is compact relative to $(X, \mathcal{B}_\beta, \mu, T)$.*
3. *For each limit ordinal $\lambda \leq \alpha$, $\mathcal{B}_\lambda = \bigcup_{\beta < \lambda} \mathcal{B}_\beta$.*
4. *$(X, \mathcal{B}, \mu, T)$ is weak mixing relative to $(X, \mathcal{B}_\alpha, \mu, T)$.*

This is the ergodic analogue of the number-theoretic result discussed above in the quote by Tao. In essence, it says that *any* measure-preserving system, irrespective of its given behavior, can be decomposed in a very precise way, viz., in terms of structured (compact) and random (weak mixing) components. Since it is easy to show that 3.2 holds in each of these cases, we can show that it holds for *any* measure-preserving system, and thus, by Furstenberg Correspondence, that Szemerédi's theorem holds.[68]

Thus, the ergodic approach to Szemerédi's theorem yields a highly perspicuous proof structure, which, in every case considered, will us tell *why* a system (and thus subset of integers) has the recurrence pattern (arithmetic progression) it does:[69]

1. **Correspondence Stage**
    (a) Convert Szemerédi's theorem into a theorem about recurrence in a measure-preserving system, i.e., Furstenberg Multiple Recurrence (Theorem 3.3).
    (b) Prove that these theorems are equivalent (or simply that Furstenberg Multiple Recurrence implies Szemerédi's theorem).
2. **Structure Theorem Stage**
    (a) Prove Furstenberg Multiple Recurrence.
    (b) This must be done for any measure-preserving system for arbitrary k.
    (c) Classify each system as either structured (compact), random (weak-mixing), or neither. If the system is structured or random, we are done.
    (d) If the system is neither, use extensions to construct a (transfinite) tower via the Furstenberg Structure Theorem (Theorem 3.5), and eventually reduce to either the compact or weak mixing case. Q.E.D.

3.2. The Structure Theorem as "Reason Why". I claim that the Furstenberg Structure Theorem provides the reason why Szemerédi's theorem holds. But what,

---

[67] A *factor* of some measure-preserving system $(X, \mathcal{B}, \mu, T)$ is a T-invariant sub-algebra of $\mathcal{B}$. That is, some σ-algebra $\mathcal{B}' \subseteq \mathcal{B}$ such that $TE, T^{-1}E \in \mathcal{B}'$ for any $E \in \mathcal{B}'$. More generally, we consider a factor to be a subsystem $X' := (X, \mathcal{B}', \mu, T)$. A factor is *trivial* if the measure of any $E \in \mathcal{B}'$ is either 0 or 1. Factors are a special case of the more general notion of extension map discussed above. The particular details of this are not terribly germane for our purposes.

[68] We must also show that Equation 3.2 lifts through limits as well as compact and weak mixing extensions, but this is all relatively routine.

[69] See the following section for a more thorough discussion of how the Structure Theorem is crucial to seeing why the result holds.

precisely, do I mean by this? Recall that, on the number-theoretic version of Szemerédi's theorem, we consider arbitrary subsets $A \subset \mathbb{Z}$ and show that they contain arbitrarily long arithmetic progressions. Unsurprisingly, given the lack of explicit constraints on A, one can find such arithmetic progressions in various subsets for very different reasons.[70] For example, consider some *random* subset A of $\mathbb{Z}$, where this means each integer n in A has an independent probability of δ with $0 < \delta < 1$. It is now quite simple to show that A in fact has infinitely many arithmetic progressions of length k because any such progression will have a probability of δ^k of lying in A.[71] On the other hand, were we to consider some "structured" set, we would get the same answer for an entirely different reason. For example, consider the *Bohr set* $\{n \in \mathbb{Z} : ||n\alpha||_{\mathbb{R}/\mathbb{Z}} \leq \delta\}$ with δ as above, $\alpha \in \mathbb{R}$, and $|| \cdot ||$ yielding the distance between the argument and the nearest integer. For any α, each αn can be made arbitrarily close to n, and so each n is correlated with periods of α. The sequence of such periods will then just be the arithmetic progression we seek.

As I showed above, the ergodic situation precisely mirrors the number-theoretic situation. Namely, we obtain Furstenberg Multiple Recurrence (Equation 3.2) in an arbitrary measure preserving system because systems exhibit either sufficient randomness (weak-mixing) or sufficient structure (compactness). In the weak-mixing case, on average, the events $T^n E, T^{2n} E, \ldots, T^{(k-1)n} E$ are uncorrelated such that the measure of the set in question in 3.2 is approximately $\mu(E)^k$ fairly often (thus satisfying 3.2). As in the random number-theoretic case, since the systems under consideration are sufficiently mixing (random), sooner or later we will find the desired recurrence property "by dumb luck" (see [7], Section 4, for a nice discussion of the structure theorem). We will also find the recurrence property in sufficiently structured systems (compact) because these systems will have events $T^n E$ recur at regular intervals, i.e., the intersection of E and $T^n E$ over k-many iterations of T^n will be non-empty. Again, this is analogous to the number-theoretic case described above: sufficient structure in the original set (or system) will guarantee the existence of the desired arithmetic progression (or recurrence property).

The moral here is that we can find multiple recurrence patterns in random systems, structured systems, and various systems in between; however, the reason we can do this always turns upon the classification of the system itself. This would seem to dash any hopes of a general reason for a result like Furstenberg Multiple Recurrence (or Szemerédi's theorem); namely, the proof of the theorem would have to amount to checking the existence of recurrence patterns for each X, given some classification of X. Such a task seems hopeless, and I daresay would not produce an explanation of why the theorem is true. One could not then even state the theorem as such; rather, we would have some conjunction of independent theorems. In order to avoid this situation, one would hope for a result that would allow us to group *prima facie* very different systems into one class that shares some feature F, which in turn allows us to understand why recurrence patterns occur in each X. *And this is just what the Furstenberg Structure*

---

[70] See [60] for an extended discussion.

[71] Take arbitrary k-length arithmetic progression $a, a + r, \ldots, a(k-1)r$. Each term contributes an independent probability of δ and we have k terms, thus the total probability of finding the arithmetic progression is δ^k.

Theorem[72] *provides for the ergodic equivalent of Szemerédi's theorem.* Thus, the reason why we can always find the desired recurrence pattern for any system X is that every X can be decomposed into components that we know how to handle, thereby obviating the need to classify each X and check for recurrence patterns. Now, one might reasonably ask: why do we need the ergodic setting to do this? If all proofs of Szemerédi's theorem use a Structure Theorem, why can't we run the aforementioned process for arbitrary subsets $A \subset \mathbb{Z}$ and find arbitrarily long arithmetic progressions? These questions are important and will be answered more fully in another paper where I explicitly compare both the ergodic and combinatorial proofs. Suffice it to say, for now, that the overall proof structure in the combinatorial case does not resemble that of the ergodic proof given above.[73] One cannot easily "read off" the existence of arithmetic progressions from the combinatorial structure theorem as one can do with the ergodic proof. Put differently, the structural result that provides the reason why the theorem holds is rendered evident in the ergodic case, while it is not at all evident in the combinatorial case.

One might, at this point, also complain that I have ignored essential details about the systems under consideration. One might think that the real reason why some X has the recurrence pattern it does is because of the kind of system X is, e.g., highly structured, highly random, etc., and that the kind of high-level structural result I consider abstracts from these details. The way to actually *explain* why a particular system has the patterns it does requires a careful analysis of *that* system in particular. There may be something to this complaint.[74] However, though this kind of criticism may be fairly levelled at particular high-level structural results, it does not really hold here.[75] As we have seen, the Furstenberg Structure Theorem gives an exhaustive means of characterizing each system (and set given our Correspondence Principle) without, as it were, digging into the nitty gritty details of the particular system (or set). When we apply the structure theorem to a particular system, even though the system might lie anywhere along the spectrum of structure and randomness, we will be able to classify the system precisely.

This classification does rely on our understanding the two ends of the spectrum, i.e., why a random system and why a highly structured system will have the recurrence pattern.[76] However, this is the only requisite contribution of "lower-level" or "system-specific" results, and this contribution is relatively simplistic, i.e., the structured and

---

[72] Indeed, it should be noted (again) that *every* proof of Szemerédi's theorem (combinatorial, ergodic, Fourier analytic, hypergraph) makes essential use of some structure theorem concerning how an arbitrary object can be decomposed into structured and random parts.

[73] Indeed, simply compare the diagram given by Szemerédi in Section 6 and the summary of the ergodic proof provided above.

[74] Indeed, I think this may be the mathematical analogue of Morrison's thesis that techniques of unification invariably abstract from "mechanisms" of physical systems which explain the behavior of the system. See Section 1.2.

[75] There may also be no real conflict here once we distinguish what it is that we are explaining. Namely, explaining why a particular system exhibits a particular behavior differs from explaining why all systems exhibit a particular behavior.

[76] Strictly speaking, however, this is not necessary. One can show that Furstenberg Multiple Recurrence holds for weak-mixing and compact systems because of very general results about extensions.

random extremes are quite easy to understand.[77] Once we know how to deal with the extremes and have the Structure Theorem in hand, the actual classifications of the various systems (or sets) do not really matter and certainly do not contribute to the explanation of why Furstenberg Multiple Recurrence (and thus Szemerédi's theorem) holds. We bypass examining each particular system because each system will have the pattern it does in virtue of the same reasons that compact and weak mixing systems have this pattern; however, the only reason we can assert this is because of the Furstenberg Structure Theorem. Thus, I believe the Furstenberg Structure Theorem makes very explicit the reason why Szemerédi's theorem is true, and thus the ergodic proof should be considered, in the basic sense of answering the why-question latent in the theorem, explanatory. Let me now turn to an explicit discussion of how this explanatory work can be done by such intuitively foreign concepts.

§4. Conceptual convergences and mathematical content.

4.1. Motivating the investigation of content. It is quite striking that Szemerédi's theorem, a finitary and combinatorial result, can be proved using infinitary and ergodic techniques. It is even more striking that these techniques yield an explanatory proof (or so I claim). We may then say that Szemerédi's theorem possesses an explanatory proof that is both elementally and topically impure.[78] But how is this possible? How do such intuitively different conceptual resources "get a grip on" the theorem to be proved? In this section, I will focus upon answering this question and thus explicate the relationship between explanation and topical impurity.[79]

Recall that a proof is topically pure if it draws only on what belongs to the content of the theorem or what the theorem is about. It would seem that the ergodic proof of Szemerédi's theorem is topically impure since ergodic theory is concerned with very different entities than those present in the theorem: at the most basic level of analysis, measure-preserving systems are quite different from sets of natural numbers or integers.[80] And yet, the structural fact that serves as the "reason why" Szemerédi's theorem holds is present in both the ergodic and combinatorial settings (the Furstenberg Structure Theorem and the Szemerédi Regularity Lemma, respectively). It is difficult to square all this information, and I believe we are faced with a "content-purity" dilemma.

---

[77] See my discussion of the weak-mixing and compact systems in the previous subsection as well as the random and structured sets in this section.

[78] See this distinction in the introduction.

[79] See [7, 8] in which a "metamathematical explanation" of the intervention of infinitary resources is given. This can be understood as an explanation of the intervention of elementally impure resources. My aim in this section is to provide a philosophical companion piece concerning the topical, rather than the elemental, impurity of the ergodic proof of Szemerédi's theorem.

[80] This can be easily seen by examining the additional structure that the underlying space X of a measure-preserving system possesses in comparison to a large set of integers or natural numbers (without additional constraints). It should be noted, however, that Furstenberg Multiple Recurrence requires few assumptions on either the probability space, the map T, or subset $E \subset X$. This mirrors the lack of assumptions (aside from positive density) on the set A in Szemerédi's theorem.

On the one hand, are we to say that, in some sense, Szemerédi's theorem has "implicit" or "hidden" ergodic content in virtue of the shared structural fact? This kind of suggestion has an impressive pedigree. Bourbaki claims something in this vicinity,[81] though I take the ultimate point to be more subtle:

> Where the superficial observer sees only two, or several, quite distinct theories, lending one another 'unexpected support' through the intervention of a mathematician of genius, the axiomatic method teaches us to look for the deep-lying reasons for such a discovery, to find the common ideas of these theories, buried under the accumulation of details properly belonging to each of them, to bring these ideas forward and out them in their proper light ([11], p. 223).

If, however, we follow the Bourbakiste suggestion (or some coarse version of it), then we should say that the ergodic proof of Szemerédi's theorem is topically *pure*. I take this to be a strange and unwelcome consequence. Furthermore, the notion of "implicit" or "hidden" content, as it stands, is too vague to be of service. I consider some proposals below that attempt to provide a more useful notion of "hidden" content.

On the other hand, we might retain the idea that the entities of ergodic theory are intuitively quite different from those of combinatorics (and Szemerédi's theorem), and thus the ergodic proof is topically *impure* as I have claimed. Obviously, I subscribe to this view, as I wish to argue for the association between impurity and explanation. However, merely relying upon the "first glance" or "intuitive" characterization of the content of a theorem leaves the explanatory potential of the ergodic setting itself unexplained.[82] Again, we should ask how the intuitively different ergodic resources "get a grip on" the explicitly combinatorial Szemerédi's theorem and how these new resources are then exploited to generate an explanatory proof. I believe that it is reasonable to answer the first question in terms of the *content* of Szemerédi's theorem, especially since topical purity/impurity is also analyzed in this way. Given that neither the "implicit" nor the "intuitive" analysis of mathematical content[83] sufficiently answers this question, I argue for a new notion called *structural content*, which is able to save intuitive purity/impurity ascriptions and simultaneously explicate the intervention of surprising and explanatorily rich conceptual resources. Let us begin, then, by characterizing the various types of mathematical content on offer in greater detail.

4.2. Intuitive and formal mathematical content. A natural starting point is the analysis given in [5]. Here Arana and Mancosu draw an important distinction between "informal" or "intuitive" content and "formal" or "axiomatic" content. Loosely, the former notion amounts to what someone with a casual acquaintance with mathematics would understand by some statement. This might profitably be understood as what one "grasps" upon a first reading.[84] For instance, Szemerédi's theorem is about a particular

---

81 In [16], Detlefsen and Arana (along with McLarty) interpret Bourbaki's quote as claiming that, e.g., particular arithmetical statements have topological content. Incidentally, the discussion of [16] involves another result due to Furstenberg: his topological proof of the infinitude of primes.

82 See my discussion of [5] below for the notion of "intuitive content."

83 At least as they stand. Various notions of content are discussed in detail below.

84 Thus distinguishing it from a higher grade of understanding one might obtain when cognizing an explanatory proof of the statement.

kind of regularity in sufficiently dense subsets of the integers (infinitary formulation) or in sufficiently large subsets of the natural numbers (finitary formulation). On the other hand, the formal notion of content amounts to the "inferential role of [a] statement within an axiomatic system" ([5], 327).

This distinction is drawn in the context of their discussion of Desargues' theorem, the exact statement of which does not matter for our purposes. One of the primary questions under investigation is whether the spatial proof of this (ostensibly) planar result is to be judged as impure (following Hilbert) or pure (following Michael Hallett) in light of metamathematical data from Hilbert's *Grundlagen der Geometrie*. In particular, Arana and Mancosu are concerned with articulating a notion of mathematical content that "...can support an adequate account of talk of purity in mathematical practice" ([5], 324). They conclude that only intuitive content is able do this work. I take no issue with this conclusion; indeed, my assertion that the ergodic proof of Szemerédi's theorem is (topically) impure relies upon the availability and cogency of intuitive content. However, it is worth considering whether we might utilize other, more nuanced, articulations of mathematical content in cases that warrant them. The theorem considered here is one such case: intuitive content cannot help us make sense of why it is that the ergodic setting is adept at modeling and explaining Szemerédi's theorem. I outline a third kind of content below that I believe is up to the task; however, before turning to this positive proposal, let me reflect on the less familiar notion of formal content to see whether some version of it might provide a serviceable sort of "hidden" or "implicit" content.

4.2.1. Isaacson's "hidden" higher-order concepts and formal content. Arana and Mancosu consider Michael Hallett's claim that the spatial proof of the planar Desargues' theorem is pure because the theorem itself has "tacit spatial content" (see [32]). This notion of "tacit" or "hidden" content descends from Isaacson's influential discussion of "hidden higher-order concepts" in [34]. As such, it is worth analyzing Isaacson's account in detail. His main thesis is that some mathematical truths expressible in arithmetic, e.g., Gödel sentences, contain "hidden higher-order concepts," and thus first-order Peano arithmetic (PA) is actually complete with respect to *genuinely arithmetical* sentences. By higher-order,[85] Isaacson means:

> [T]he standard usage for quantification over sets of individuals in distinction to first-order quantification over the individuals themselves. But I also mean to include in this phrase something of the notion of the infinitary, in the sense of presupposing an infinite totality [...] ([34], p. 210).

According to Isaacson, these higher-order concepts are implicit in arithmetical propositions via the coding of various syntactic properties and relations by properties and relations of the natural numbers (as in Gödel's proof of the Incompleteness Theorems). Put more precisely, a Gödel sentence *G* can be shown in PA to be equivalent to a sentence expressing (by coding) a metamathematical property of PA, e.g., Con(PA). This metamathematical property is unprovable in PA, but can be proven

---

[85] See the beginning of Section 2.2 for my explication of the the finitary–infinitary distinction which maps quite closely on to Isaacson's own.

using higher-order (in the sense above) concepts. Thus, the equivalence reveals the implicit higher-order content of *G*.

A brief digression:[86] Isaacson does not say anything further about the exact nature of the implicit higher-order content of *G*, but presumably this is gotten from the arithmetical coding of transfinite ordinals required for the proof of Con(PA). In particular, we know that a finitistically acceptable theory (e.g., PRA) augmented with a principle expressing (by coding) transfinite induction along a well-ordering of the ordinals up to a particular ordinal, ε_0, proves Con(PA).[87] This principle must, then, provide the content of Con(PA) (and thus of *G*) that is higher-order and non-arithmetical; however, the exact nature of this content is not so straightforward given that, for example, Gentzen's proof of Con(PA) does not *explicitly* rely upon transfinite ordinals. Rather, it relies upon induction along a well-ordering of *ordinal notations* (coded by natural numbers), and this well-ordering is isomorphic to the well-ordering of transfinite ordinals up to ε_0. Nonetheless, the idea seems to be that the use of this additional machinery in developing ordinal notations suffices to impart infinitary higher-order content to *G*. This is because of Isaacson's understanding of what counts as a genuinely arithmetical truth:

> [A] truth expressed in the (first-order) language of arithmetic is arithmetical just in case [(i)] its truth is directly perceivable on the basis of our (higher-order) articulation of our grasp of the structure of the natural numbers *or* [(ii)] directly perceivable from truths in the language of arithmetic which are themselves arithmetical ([34], p. 217).

Thus, for Isaacson, there is an innocuous sort of higher-order content (given by (i)), the use of which does not render truths expressible in L_{PA} non-arithmetical. For instance, if we come to perceive the truth of some φ expressible in L_{PA} from our understanding of Dedekind's second-order, categorical characterization of the natural numbers, then φ is still arithmetical. On the other hand, the Gödel sentence *G*, although expressible in L_{PA}, is rendered non-arithmetical because: (a) it is equivalent to a sentence ψ expressible in L_{PA} that encodes Con(PA); (b) the proof of Con(PA) requires induction on the well-ordering of ordinal notations for ordinals $< \varepsilon_0$. The question of *G*'s higher-order content thus reduces to that of whether this inductive principle can be justified arithmetically. Isaacson claims that it is "reasonably evident" that *G* is not arithmetical,[88] and so the answer to this question is no. I am inclined to agree with his analysis; however, I am not sure that the reasons supporting it are so evident. First, we would need to show that there is no arithmetical means by which the well-foundedness of the ordering of the codes of the ordinals $< \varepsilon_0$ could be understood. I find it plausible that this could be arithmetically justified for sufficiently small transfinite ordinals, perhaps, e.g., $\omega + \omega, \omega + \omega + \omega, \ldots, \omega \cdot \omega$. However, for the

---

[86] The uninterested reader may skip this digression as it does not affect the thesis of my paper. I merely wished to elaborate on the details of Isaacson's original proposal about "hidden" or "implicit" content.

[87] The ordinal ε_0 is the limit ordinal of the increasing sequence $\omega < \omega^\omega < \omega^{\omega^\omega} < \cdots$. Also, note that the ordinals $< \varepsilon_0$ must be written in a particular way, i.e., in Cantor normal form.

[88] At least, in the sense of (i), since it is possible that new proofs will emerge that proceed from "recognizably arithmetical" truths.

purposes of obtaining a proof of Con(PA), we must be able to perceive arithmetically the truth of induction along the well-ordering of codes of ordinals *up to ε_0*, and this is a much more imposing requirement.[89] The nature of the structure of this well-ordering seems sufficiently complicated to render it higher-order and thus non-arithmetical. I take it that something along these lines is what Isaacson means by the implicit higher-order content of G.

4.2.2. Can we utilize formal content?. According to Arana and Mancosu, the case of Desargues' theorem is analogous to that of the Gödel sentence because it can serve as a spatial incidence axiom in Hilbert's axiomatic system of the *Grundlagen*, i.e., the planar Desargues' theorem can play the same inferential role as an explicitly spatial mathematical proposition.[90] They summarize the situation in the following way:

> [I]n both cases, we have sentences whose ordinary understanding indicates that it has content of one type (e.g., arithmetical or planar), but that an analysis of the sentences' inferential roles reveals that these sentences have tacit content of another type (e.g., infinitary or spatial, respectively) ([5], p. 333).

Nonetheless, even though they believe that the Isaacson–Hallett notion of formal content makes sense for Desargues' theorem, they ultimately reject its use. And this is because appealing to formal content generates verdicts contrary to mathematical practice; most pressingly, purity ascriptions become trivial. Related considerations and worries will appear below. However, at present we only consider whether it is intelligible to say that Szemerédi's theorem has formal content and, if so, whether formal content can help to explicate the relationship between impurity and explanation.

There are, certainly, some similarities between our case study and the phenomenon of hidden content at work in Gödel sentences and Desargues' theorem. Our ordinary or intuitive understanding of Szemerédi's theorem indicates that it has finitary, combinatorial content; however, it is equivalent to an ostensibly infinitary result (Furstenberg Multiple Recurrence), and its ergodic proof makes essential use of a transfinite construction[91] (the tower of extensions in the Furstenberg Structure Theorem). This initial assessment of the situation suggests that the notion of formal content may be of use in explicating this surprising confluence of mathematical resources. However, closer examination reveals that formal content is not really available to us. Consider the case of the Gödel sentence G. Let us write the formal content of G as:

$$\mathcal{F}_G := \{\psi : \vdash_{\mathsf{PA}} G \leftrightarrow \psi\}. \tag{4.1}$$

---

[89] It is clear that the innocuous higher-order content (Dedekind's categorical characterization of the natural numbers) cannot justify the resources needed to perceive the truth of Con(PA). That is, there is no way to proceed from Dedekind's characterization to the principle expressing transfinite induction upon to ε_0.

[90] See [5], Section 4 for more details.

[91] Of course, we can ask *how much* of this construction do we require? Is the full strength of the Furstenberg Structure Theorem needed? See [8] for a negative answer to this latter question. There are many fascinating details about the axiomatic strength of the ergodic proof which I suppress here.

That is, the formal content of G is the class of statements such that each statement is provably equivalent to G in PA. Thus, as it should be, some ψ expressing by coding $\mathrm{Con}(\mathrm{PA})$ is included in the formal content of G. However, when we think of formal content in the above fashion, there is a rather restricted set of circumstances in which this will make sense, viz., when we have an independence result. In the case of G, while we can prove $G \leftrightarrow \psi$ in PA, we cannot prove the property that ψ expresses, i.e., $\mathrm{Con}(\mathrm{PA})$, in PA by Gödel's Second Incompleteness Theorem. Thus, we have the requisite independence. In the case of Szemerédi's theorem (SZ), we could try to write its formal content as:

$$\mathcal{F}_{\mathsf{SZ}} := \{\psi : \vdash_{\mathsf{ZFC}} \mathsf{SZ} \leftrightarrow \psi\}. \tag{4.2}$$

Letting some ψ be Furstenberg Multiple Recurrence, we would register the "hidden" infinitary ergodic content of Szemerédi's theorem. However, Furstenberg Multiple Recurrence is not independent of ZFC, the system in which we prove the equivalence. This has the effect of saying that Szemerédi's theorem and Furstenberg Multiple Recurrence have the *same* formal content; indeed, all theorems of ZFC would then have the same formal content, rendering this notion (as it stands) entirely trivial and, *a fortiori*, entirely unhelpful in illuminating the relationship between impurity and explanation.

Another way of seeing the unavailability of formal content in our case is via the absence of a coding phenomenon. According to Isaacson, "The relationship of coding constitutes a rigid link between the arithmetical and the higher-order truths, which pulls the ostensibly arithmetical truth up into the higher-order" ([34], 221). There is no such mechanism at work in the Furstenberg Correspondence Principle.[92] Rather, we have an identification of sets and functions in the combinatorial and ergodic contexts, along with a clever choice of "unconventional" dynamical system that effects this identification.[93] This should be thought of as a "modeling" of the combinatorial phenomenon in the ergodic setting, rather than indicating the presence of coded or "hidden" formal content.[94]

Of course, these considerations need not rule out the possibility of using formal content to explicate Furstenberg Correspondence in the future. This would depend upon a reverse mathematical analysis of both Szemerédi's theorem and Furstenberg Multiple Recurrence and the existence of an independence result.[95] That is, we would need to show that some system T proves Szemerédi's theorem and its equivalence with Furstenberg Multiple Recurrence but not Multiple Recurrence itself. It is very likely that Szemerédi's theorem can be proved in RCA_0 and perhaps even weaker systems, although this would require some rather tedious work to verify. Furthermore, the metamathematical analysis of the infinitary content of Furstenberg Correspondence in [7], Section 5, indicates that it is reinterpretable in computational or combinatorial

---

92 Arana and Mancosu note that there is no *explicit* coding in the case of Desargues' theorem either but surmise that it is given by the "...algebra of segments permitted by Desargues' theorem in the presence of axioms I 1-2, II, and III, or more directly the alleged spatial content it inherits from its stereometrical proof" ([5], p. 333).

93 I have not discussed this in my proof sketch above; see [24] for details.

94 One might also say that we adopt a new sort of perspective on the combinatorial results by considering them *qua* ergodic.

95 See [57] for the canonical presentation of the reverse mathematics program.

terms. Thus, given this metamathematical data, it would appear that the independence result required for formal content to make sense is unlikely to materialize.

Finally, it is worth considering whether a looser notion of hidden higher-order content might be of interest. This would then no longer involve Arana and Mancosu's gloss of hidden higher-order content as *formal content*, and thus the need for an independence result would no longer be present. Isaacson himself suggests something along these lines when dealing with "in principle" provability. That is, there might be cases where a theorem is provable in PA, but this proof would be "infeasibly long" and thus "the higher-order perspective is essential for *actual* conviction as to the truth of an arithmetically expressed sentence" ([34], p. 221). I am rather sympathetic to the idea that various statements in some L_T could be said to have higher-order content in light of the unsurveyability of their proofs[96] in the system T.[97] It is implausible that Szemerédi's theorem has hidden higher-order content in this sense: its proof is very *difficult* but is by no means unsurveyable. However, something like, say, Fermat's Last Theorem (FLT) seems a good candidate for an ascription of "loose" higher-order content. Few seem to doubt that it is "in principle" provable in some (higher-order) non-conservative extension of PA; however, actually producing and understanding such a proof appears to be nearly impossible (see [47]). It is then reasonable to ascribe *some sort* of hidden higher-order content to FLT but not to Szemerédi's theorem.

4.3. A further refinement of mathematical content. Thus, though the notion of formal content and looser versions of "hidden" higher-order content may be useful in particular circumstances, they cannot help us to understand the case at hand. Should we then evaluate Szemerédi's theorem in terms of intuitive content alone? Unfortunately, as I have indicated, this articulation of content fails to capture crucial data about the theorem, data that would be desirable to have for philosophical discussions relating content, purity, explanation, mathematical evidence, and likely other topics. In particular, we should like to say something about the relationship between the combinatorial and ergodic settings in light of both the equivalence between Szemerédi's theorem and Furstenberg Multiple Recurrence and the epistemic dividends generated by the ergodic proof. I find it quite reasonable that this connection be captured in our understanding of the content of Szemerédi's theorem. In short, then, intuitive content fails to capture mathematical information of interest, while formal content is unavailable to us and, even if it were, is neglectful of important epistemic distinctions.[98] Is there then another, "intermediate" notion of content that might make sense of the

---

96 Isaacson notes that this looser notion of higher-order content would not affect his thesis that PA is complete with respect to genuinely arithmetical truths because any expansion in the class of statements possessing higher-order content would be accompanied by a narrowing of the domain of the arithmetically true.

97 Indeed, this kind of thought is closely related to my claims in another paper that, even if infinitary resources are not proof-theoretically *required*, they are still in some sense necessary for intelligible mathematics (or, in Isaacson's words, necessary for "actual conviction"). I discuss Fermat's Last Theorem in greater detail there.

98 For instance, Arana and Mancosu reject formal content as useful in making purity ascriptions because it has the following consequences: (i) The very question of content cannot be articulated using only formal content (as I have noted at the outset); (ii) Someone without beliefs concerning space could not understand Desargues' theorem; (iii) Purity ascriptions would be entirely trivial as every theorem would have a pure proof; (iv) The

ergodic proof of Szemerédi's theorem? I believe that there is and attempt to use the somewhat sketchy suggestions of Bourbaki to develop this idea. Call this third sort of content *structural content*.

Bourbaki describes the "axiomatic method" as a "systematic study of the relations existing between different mathematical theories," which takes as its central concept that of the mathematical structure ([11], 222). Consider Bourbaki's description of a mathematical structure:

> The common character of the different concepts designated by this generic name [mathematical structure], is that they can be applied to sets of elements whose nature has not been specified; to define a structure, one takes as given one or several relations, into which the elements enter [...] [T]hen one postulates that the given relation, or relations, satisfy certain conditions (which are explicitly stated and which are the axioms of the structure under consideration) (*ibid.*, p. 226).

I propose that the structural content of a theorem is the instantiation of a particular fundamental mathematical structure by the entities intuitively involved in the statement. Let me fix intuitions with a simple example: the realization that $(\mathbb{R}, +)$ is a particular instance of a group structure[99] contributes additional content to a theorem involving addition on the real numbers. This content cannot be what we are calling intuitive: it seems doubtful that a mathematical novice proving a theorem about $(\mathbb{R}, +)$ that is not explicitly group-theoretic would grasp the group axioms. Nor would a theorem involving $(\mathbb{R}, +)$ gain a contribution of strictly formal content from the realization that $(\mathbb{R}, +)$ is the instantiation of a fundamental structure; the understanding that $(\mathbb{R}, +)$ is a group does not involve seeing that it can *serve* as an axiom in a system but rather the understanding that $(\mathbb{R}, +)$ *satisfies* particular axioms. Furthermore, this realization will induce a relational fact about the particular entity in question, $(\mathbb{R}, +)$, and other groups, i.e., the fact that they all instantiate the group structure. This fact contributes content that is of a much more global character than intuitive content, while not being merely inferential. We might then fruitfully compare various instances of group structures once we realize that they can all be identified as such.

Bourbaki is careful to note that this description of the mathematical universe via structures is "...schematic, and idealized as well as frozen" (*ibid.*, p. 229). As such, it is open to revision, and, in particular "[the] definition of structures is not sufficiently general" (*ibid.*, footnote on p. 226). I want to suggest that we can proceed by taking a mathematical structure to be a somewhat more flexible notion than a concept defined by explicit axioms. Of course, once we broaden the notion of mathematical structure to include more than explicit axiomatizations, some further attempt at characterization must be made especially if we wish to avoid pointless "gerrymandered" properties. A natural suggestion is that we appeal to compositional facts about mathematical

---

content of statements would be radically contextualized, e.g., it would depend entirely upon the axiomatic context.

[99] Where this structure is defined by the group axioms: closure, associativity, existence of identity and inverse.

objects: for some object X in (conceptual) domain D write

$$X \text{ can be identified as } Y \sim Z \tag{4.3}$$

for more restricted objects Y, Z in the same domain under appropriate relation $\sim$ (this may be a cross product, direct product, set-theoretic union, etc.). It is not difficult to find such results, especially in algebraic contexts. Indeed, there are vast swathes of mathematics concerned with "structure theorems": structure theory of countable Abelian groups, of semisimple Lie algebras, of group schemes of finite type, etc. I mention these other examples to convince my reader that the compositional facts I advert to are deeply embedded in mathematical practice, and, as such, provide a reasonable starting point for my intermediate notion of content.[100]

These compositional facts are not restricted to algebraic contexts.[101] We have seen that subsets of integers of positive density can be decomposed in such a fashion by appeal to a Structure Theorem, which expresses the dichotomy between structure and randomness.[102] Thus, according to my content proposal, Szemerédi's theorem, simply by virtue of the fact that it involves sufficiently dense subsets of $\mathbb{Z}$, has as part of its content the dichotomy between structure and randomness. Even though the theorem is not obviously about this dichotomy, as an intuitive reading or the "verbal form" of the mathematical proposition[103] would suggest, I believe we should consider the instantiation of this dichotomy to be included in its content. However, and here is the crucial point, this structural fact about how the objects in question may be decomposed need not be (and is not) endemic to the integers or the naturals alone. One can similarly see the dichotomy at work in the ergodic setting: measure-preserving systems also exhibit this structural fact as expressed by the Furstenberg Structure Theorem. We might think of the different mathematical objects in question (sets of integers, measure-preserving systems) as multiple realizations of this same high-level structural fact.[104] Thus, by appealing to structural content, we allow that intuitively different mathematical domains retain some degree of conceptual independence, facilitating

---

[100] I do not claim such structural results function in the same way as the dichotomy between structure and randomness in Szemerédi's theorem. Such a claim would inevitably involve a careful analysis of particular algebraic structure theorems.

[101] For instance, Avigad remarks, "This opposition [the dichotomy between structure and randomness; more precisely, the dichotomy between compact and weak mixing behavior] is fundamental in analysis: a system can be rigid, or chaotic; a channel can carry signal, or noise" ([7], p. 68).

[102] See Tao's quote in Section 3.1.

[103] Wittgenstein raises a similar question in his *Remarks on the Foundations of Mathematics*. I suppose I am in broad agreement with his claim that: "[T]he understanding of a mathematical proposition is not guaranteed by its verbal form, as is the case with most non-mathematical propositions" ([66], IV, 25). Unfortunately, the way in which he elaborates on this is hardly serviceable. He proposes that the sense of a mathematical statement is gotten from its proof. But how can this be? How can one even set out to prove a statement that has no initial sense? This embroils us in a sort of Meno problem, the only escape from which requires that we know *something* initially. I propose that this will be the "intuitive content" of a mathematical statement.

[104] This idea is common in the philosophy of science literature. Multiple realizability is, more or less, the idea that there can be heterogeneous "realizers" of "upper-level" properties and generalizations. That is, these generalizations characterize the same behavior in physically distinct systems.

various practical distinctions like purity ascriptions, but we gain a principled way to talk about surprising confluences of such domains. Both the ergodic and number-theoretic settings exhibit the structural dichotomy in question, i.e., they both encode the same sort of information (the confluence), but each context obviously involves different objects that serve to made the dichotomy precise in different ways (the independence).

A little more explicitly: the proofs of Szemerédi's theorem (T_C; Theorem 2.5) and Furstenberg Multiple Recurrence (T_E; Theorem 3.3) rely on their respective Structure Theorems, the Szemerédi Regularity Lemma[105] (S_C; Theorem 6.1), and the Furstenberg Structure Theorem (S_E; Theorem 3.5). Thus, we have a class of theorems $\mathcal{T}$, of which T_C, T_E are members, defined by the fact that all theorems in the class rely on Structure Theorems (theorems showing how a particular entity can be decomposed into a structured and random part) for their proof. The presence of Structure Theorems in these proofs then indicates that there is a general structural fact in the offing, i.e., the dichotomy between structure and randomness. And so, we have found an incredibly important higher level property of the theorems in $\mathcal{T}$, a property that is not made explicit by the intuitive content of the theorems. Crucially, the fact that $T_C, T_E \in \mathcal{T}$ allows us to make sense of the surprising intervention of the Furstenberg Structure Theorem in the proof of Szemerédi's theorem and makes the ergodic context suitable to prove a number theoretic result. This is because the objects manipulated in each context (subsets of integers and measure-preserving systems) can be decomposed into structured and random components.[106] The fundamental dichotomy between structure and randomness allows us to make sense of the confluence of these *prima facie* very different domains, and its clear expression in the ergodic context will help to render the ergodic proof of Szemerédi's theorem explanatory.

4.4. An objection and two interpretations of my view.   Let me respond to a natural objection. One might think that in offering this notion of structural content I have tried to have my cake and eat it too. I have tried to provide a notion of content that retains intuitive ascriptions of purity/impurity and also explicates the intervention of impure resources. This has been done, more or less, by carving out a genus of mathematical results, a sort of mathematical natural kind if you will, under which both Szemerédi's theorem and Furstenberg Multiple Recurrence fall.[107] However, one might think that this then requires saying that the ergodic proof of Szemerédi's theorem is pure, thereby impaling us on the first horn of the content-purity dilemma.[108] This objection would be analogous to that leveled by Arana and Mancosu against proponents of formal content:[109] privileging formal content "...threatens to trivialize purity, in making it the case that *every* theorem has a pure proof, when the content of that theorem is

---

[105] I have not explicitly discussed this theorem because I have considered only the ergodic proof in this paper. I compare the combinatorial and ergodic proofs of Szemerédi's theorem and discuss the Regularity Lemma at some length in another paper.

[106] Of course, here we do not have axioms defining a structure which particular entities instantiate. Nonetheless, at least in the case under consideration, both "structure" and "randomness" can be made perfectly precise in each context, viz., combinatorial and ergodic.

[107] As well as the analogous results in Fourier analysis and hypergraph theory and special cases of all these theorems. Thus, this is quite a robust genus.

[108] See Section 4.1.

[109] This easily translates into Aristotelian terms as well. We have not in fact "crossed genera" when we utilize structural content. The issue I indicate here thus precisely mirrors that in the scholarship on Aristotle's theory of demonstration, viz., how are we to understand the

fully understood" ([5], p. 336). Similarly, one might say that every theorem has a pure proof once investigators have dug "deeply enough" and uncovered the requisite structural similarities. Another way of putting this objection would be to appeal to the topical content of a theorem and observe the purity metric it induces. For example, under my construal of structural content, both Szemerédi's theorem and Furstenberg Multiple Recurrence are *topically close* insofar as both subsets of integers and measure-preserving systems instantiate the dichotomy between structure and randomness. This topical closeness then indicates that the ergodic resources used to prove Szemerédi's theorem do not involve an appeal to impure techniques.

In response, I am inclined to restrict the ways in which the topical content of a theorem induces a purity constraint.[110] If we are interested in philosophically explicating mathematics as practiced, then the purity constraints we delineate should map closely onto the activity of mathematicians. There is a good deal of evidence that, when speaking of a "pure" proof, mathematicians are operating with an intuitive notion of content.[111] Thus, intuitive content is what philosophers should appeal to in making claims about the purity or impurity of proofs. Nonetheless, I believe the arguments provided above support the existence of structural content. Structural content is intelligibly part of what a theorem is about precisely because the entities *intuitively* involved in the theorem instantiate particular structural features. However, we should block the move from the presence of structural content to a structural purity constraint because the notion of purity generated is not acknowledged by the mathematical community and, like formal content, would trivialize ascriptions of purity made in practice.

Furthermore, from a strictly conceptual perspective (independent of a methodological interest to generate philosophical theses generally consonant with mathematical practice), structural content enjoys a crucial advantage over formal content: it is able to coexist with intuitive content. If both notions were not simultaneously available to us, then we would be in no better a position than the proponent of formal content. Structural content, however, keeps the original mathematical concepts in view, whereas formal content ignores these concepts altogether by privileging the inferential role of a statement as the determinant of the statement's content. Indeed, the ascription of structural content *requires* that we recognize the intuitive content of a theorem, for it is the entities intuitively recognized that instantiate the structural features of interest. We must initially acknowledge that two domains of mathematics are ostensibly unrelated but interact in surprising ways. Only then can we acknowledge and begin to uncover the "deep-lying reasons" for this interaction, reasons which are not about only one domain or the other but rather encompass both. Thus, we can retain a purity constraint generated by intuitive content: ergodic theory and combinatorics involve intuitively different concepts and movement between these domains should be considered an instance of impurity. However, these domains fall within a larger structural genus,

---

      genus in question that determines the purity constraint? Again, see [58] for a survey of some answers to this question.

[110] Thus, the distance metric given by the topical content of a theorem does not necessarily generate a purity constraint in an obvious way. One might, on the other hand, simply say that structural content is a distinct notion of content from the topical kind, but this seems quite counter-intuitive to me. I believe it is correct to say that both Szemerédi's theorem and Furstenberg Multiple Recurrence are, in some sense, *about* the dichotomy in question.

[111] See the discussion of intuitive content in [5].

which elucidates an important commonality between them without obliterating their intuitive differences (as is done by formal content). We can then comprehend the possibility of these domains interacting in an intelligible way given their membership in a common genus.

I have been quite cautious in my description of structural content. In particular, I have been arguing for a view that ergodic theory and combinatorics are genuinely distinct mathematical domains[112] but that particular propositions involving entities from each domain share structural content. As I said above, this has been done to honor mathematical practice and various intuitions and epistemic virtues derived from it. One way to put the point is that the more abstract structural content "follows" from our grasp of the "given" intuitive content. That is, we must grasp (in a relatively brute sense) what the theorem in question is about and only then can further structural facts be brought to light. Now, as I have repeatedly stressed, not any structural fact will do, e.g., inferential roles and the associated construal of content will be rather useless for understanding explanation and other epistemic virtues. Thus, it is instructive to think of all this in terms of conceptual distance: in order to make sense of what explanation in mathematics looks like, we cannot go "too far" past the intuitive content of a theorem (as formal content does). Nonetheless, I have claimed that we must go somewhat beyond it: explanatory facts are not often surface-level but require difficult stages of reflection and abstraction. This is quite reasonable because it is natural to think that the cognitive state we are in when we understand "the why" is of a much higher grade than that which we are in when we grasp the statement of the theorem or have a merely cogent (non-explanatory) proof.[113]

However, the more metaphysically adventurous may be inclined to press a little here. If the connection between Szemerédi's theorem and Furstenberg Multiple Recurrence is as robust as I claim, why retain the initial intuitive distinction? Why not, once we have ascended to the "reason why," throw away the ladder? Why not remove the linguistic veil and see the world aright? I must admit that I am not entirely sure what to think here. It is unreasonable to assimilate ergodic theory and combinatorics *tout court*; the connective tissue given by structural content is not always present. However, perhaps all theorems expressing the dichotomy between structure and randomness are, in some very strong sense, the same thing. What would the consequences of such a view be? First, obviously, we would lose all ability to make purity/impurity ascriptions. But I would imagine the adventurous reader would be unbothered by this: purity/impurity is not the interesting thing here, and all such claims should be thrown away. Second, such a strong metaphysical assertion would ripple throughout mathematical practice in different ways, e.g., new results could be absorbed into the genus we have uncovered, intuitively distinct domains would no longer appear such, etc. But, again, I do not see why the adventurous reader would care. Ultimately, then, the preferred interpretation likely reduces to one's view of the role of philosophy: should it be the queen of the sciences and not beholden to the practice of other disciplines? Or should it serve

---

[112] As with most concepts, the boundaries are relatively fuzzy. However, one can intelligibly pick out what the putative entities of interest for each domain are, and these appear quite different.

[113] Aristotle continues to be instructive as he believes the haver of ἐπιστήμη (epistēmē) is in a very privileged cognitive state.

a more interpretive role, explicating the methods and concepts of other disciplines without upsetting the scenery too much? I leave this to the tastes of my reader.

4.5. Summary and the prospect of generalization. Let me summarize the dialectic thus far. The equivalence of Szemerédi's theorem and Furstenberg Multiple Recurrence is surprising and may suggest that Szemerédi's theorem has hidden higher-order content: in this case, hidden infinitary and ergodic content. That is, in some sense, Szemerédi's theorem is about both subsets of integers and infinitary measure-preserving systems. However, a consequence of this view is that we must call the ergodic proof pure. In order to better understand whether some version of this "hidden" content view could be made intelligible, I examined Arana and Mancosu's discussion of intuitive and formal content. We saw that the notion of formal content operative in other discussions of hidden higher-order content (Isaacson and Hallett) is unavailable to us and agreed with the argument of [5] that intuitive content is essential for making purity ascriptions. Nonetheless, intuitive content fails to explicate the relationship between combinatorics and ergodic theory in the ergodic proof of Szemerédi's theorem.

Thus, I have tried to excavate a third kind of mathematical content: structural content. The ultimate point of this third kind is that it occupies the middle ground between intuitive and formal. Intuitive content is important and useful in mathematical practice, yet it does not help us to make sense of surprising convergences of mathematical results; formal content is useful only in very restricted contexts and may trivialize important epistemic distinctions. Structural content, on the other hand, is slightly coarser than intuitive content but still occurs within mathematical practice. I have described it as the instantiation of particular structural facts by intuitively recognized entities that induce higher-level dependencies between theorems, linking them in a conceptual nexus that crosses over intuitively distinct domains. Of course, unrestricted appeal to *any* property shared by various theorems (e.g., all theorems involving prime $p = 5$) will not be useful; this may be a case of "gerrymandering" mathematical properties. I have tried to avoid such an issue by suggesting an important class of compositional properties. In our case, it is clear that both Szemerédi's theorem and Furstenberg Multiple Recurrence require structure theorems in their respective proofs, i.e., the instantiation of the dichotomy between structure and randomness is apparent. Thus, the ascription of structural content is reasonable. Finally, I noted that my proposal may be interpreted in either a metaphysically weak or strong sense, and if one opts for the strong reading, then much of the above discussion about "saving the appearances" of mathematical practice becomes otiose.

Let me also note that I believe my content proposal may be generalized. Of course, this will require similar analyses of theorems proved via intuitively impure techniques. I will not undertake this here but would like to mention an exemplary case consistent with my philosophical discussion: the complex analytic (impure) and arithmetical (pure) proofs of the Prime Number Theorem.[114] The Prime Number Theorem (PNT) asserts that $\pi(x) \sim x/\log(x)$ as $x \to \infty$ where $\pi(x)$ counts the number of primes $p \leq x$ (that is, the number of primes up to integer x is approximately $x/\log(x)$ as $x \to \infty$). This was independently proved by Hadamard and de la Vallée Poussin in 1896 via appeal to complex function theory (see [14, 30]). Indeed, they showed that the PNT is in some sense equivalent to the non-existence of zeroes of the Riemann

---

[114] This has been extensively analyzed by Arana. See, for instance, [4].

zeta function $\zeta(s)$ in complex variable $s = \sigma + it$ for $\mathrm{Re}(s) = 1$. It was long thought that an, "'elementary' proof of the PNT, not depending on analytical ideas remote from the problem itself," would be desirable ([33], p. 651). This was achieved, again independently, by Selberg and Erdős whose analyses were entirely elementary and arithmetical. Many of the details are not germane to our discussion here; however, it is crucial to note the following structural similarity between the complex-analytic and arithmetical proofs as brought out by Ingham's incisive review of the Selberg and Erdős papers. Ingham extracts four analytical properties of $f :=- \zeta'/\zeta$ which "embody the essential analytical fact on which previous proofs of the PNT [impure, complex-analytic ones] have been based." He goes on to note that

> What Selberg and Erdős do is to deduce the PNT directly from the *arithmetical counterparts* of (i), (iii), and (iv), without the explicit intervention of the analytical fact. In principle this opens up the possibility of a new approach, in which the old logical arrangement is reversed and the analytical properties of $\zeta(s)$ are deduced from the arithmetical properties of the sequence of primes ([33], p. 654, emphasis my own).

Arana reads this as indicating the presence of implicit complex-analytic content in explicitly arithmetical statements (see [4]). My position should, by now, be quite predictable. What we have instead is the instantiation of higher-level structural properties by complex-analytic functions and prime numbers, respectively. Selberg and Erdős eliminate essential appeal to strictly complex-analytic content and pass to these structural features. Thus, we need not relinquish the intuitive impurity of the complex-analytic proof of PNT by appeal to mysterious "hidden" content; we have instead the presence of structural content.[115]

Finally, this discussion of content serves as a first and very important step in understanding how it is possible for impure proofs to be explanatory. The presence of structural content also helps to disarm the two conceptual reasons provided for thinking that impurity and explanation may come apart.[116] I believe that my notion of structural content provides one reason to think that these worries may not always be apposite. That is, the presence of structural content will help to ensure both that the intended why-question is being answered and that the data relevant to answering this question is preserved in a detour through impure resources. If one can show that particular structural facts crucial to the proof of a theorem T in context X also occur in context Y, then it is entirely reasonable to think that a proof of T in context Y will be a candidate explanatory proof. This is precisely what happens in the ergodic proof of Szemerédi's theorem. We see the instantiation of the dichotomy between structure and randomness in both the number-theoretic and ergodic contexts but because of the greater conceptual clarity in the ergodic context, which occurs in part because the infinitary nature of the objects smooths out many inessential details, the impure proof is explanatory, while the pure proof is not. Let me now turn to this.[117]

---

[115] This is not to say that the presence of these higher-level structural facts is not mysterious in itself.

[116] In particular, see Section 1.2, p. 8.

[117] A much more detailed discussion is provided in another paper.

§5. Conclusion: explanatory proofs via simplicity and unification. As I noted at the outset, in order to be classified as explanatory, a proof must (i) provide the putative "reason why" *and* (ii) show how this reason manifests itself *qua* explanatory fact. This somewhat vague (though philosophically important) idea can now be made more precise by examining the details of our case: first, an explanatory proof of Szemerédi's theorem must tell us why arbitrarily long arithmetic progressions occur in any sufficiently large set of integers. We have seen that the reason for this is the presence of the dichotomy between structure and randomness: any sufficiently large subset of integers can be analyzed into well-behaved/structured and random components, and any combination of these components will generate the arithmetic progressions we seek. Thus, the dichotomy provides an answer to our why-question, and, crucially, this fact occurs in both the combinatorial and ergodic settings thereby licensing the ascription of shared structural content.

However, we must now analyze how the dichotomy manifests in various proofs of Szemerédi's theorem. Do we say that both the pure (combinatorial) and impure (ergodic) proofs are explanatory in virtue of their both registering the requisite dichotomy? I believe that the answer is quite clearly no. Some of Tao's remarks on the pure proof will be of use:

> Szemerédi's original proof of this theorem is a remarkably intricate piece of combinatorial reasoning. Most proofs of theorems in mathematics–even long and difficult ones–generally come with a reasonably compact "high-level" overview, in which the proof is (conceptually, at least) broken down into simpler pieces. There may well be technical difficulties in formulating and then proving each of the component pieces, and then in fitting the pieces together, but usually the "big picture" is reasonably clear. [...] In contrast, the pieces of Szemerédi's proof are highly interlocking, particularly with regard to all the epsilon-type parameters involved; it takes quite a bit of notational setup and foundational lemmas *before the key steps of the proof can even be stated, let alone proved* (Blog post 3/23/2012; https://terrytao.wordpress.com/2012/03/23/some-ingredients-in-szemeredis-proof-of-szemeredis-theorem/; emphasis my own).

Thus, given its intricacy and "global" opacity, the pure combinatorial proof fails to yield any sort of understanding as to *why* Szemerédi's theorem is true, and so fails to be explanatory. In particular, we are not able to locate the structural and explanatory fact *qua* explanation internal to the nexus of reasoning that constitutes the proof. Somewhat ironically, the incredible combinatorial ingenuity required to establish the *truth* of the theorem obscures the *reason why* it is true. This is only a coarse summary of comparative work I perform in another paper but should suffice to indicate why the pure proof, despite utilizing the dichotomy between structure and randomness, fails to be explanatory.

Now consider the ergodic proof of Szemerédi's theorem. I have claimed that the ergodic setting is "appropriate" for providing an explanatory proof in virtue of the fact that it shares structural content with the combinatorial setting. Of course, this does not yet yield an explanation because the ergodic proof might be just as opaque as the combinatorial. This is not, however, the case (my pithy summary of the proof

at the end of Section 3.1 should lend some support for this claim). In particular, the impure, ergodic setting, supported by shared structural content, leads both to greater simplicity and unification. These are, of course, two very familiar epistemic virtues, and it would be absurd to claim any originality in isolating them as important for explanation. However, as is the case with many "familiar" notions in philosophy, we must be very careful in explicating what is meant by both; again, I will take this up in another paper. Let the following brief summary suffice for now.

First, the ergodic proof of Szemerédi's theorem is simpler in the sense that its global proof structure is much more perspicuous than that of the combinatorial proof. Very roughly, the combinatorial proof proceeds as follows:[118] construct a massive "generalized" arithmetic progression of suitably large rank,[119] arrange it so that it is "saturated" by elements of A, and then carefully winnow it down until one arrives at the desired k-term arithmetic progression (rank one) in given A. This winnowing process is particularly difficult because one must show various "nice" combinatorial conditions hold for particular indices, but once one moves to a lower index, the condition proved for larger indices might fail to hold. All of this renders the pure proof highly non-linear; Tao remarks that it resembles a huge "Tower of Hanoi" construction. On the other hand, the ergodic proof can be set forth in a linear and informative fashion that makes the "big picture" evident.[120] This happens in the ergodic setting, in part, because of its infinitary nature: the delicate combinatorial manipulations are abstracted away but without the loss of data crucial for providing an explanation. However, once more, I wish to note that the axiomatic strength of the ergodic setting is not the main point here. Rather, it is a very helpful side effect of the fact that ergodic concepts can be brought to bear on combinatorial concepts in virtue of the structural content shared between these domains.

Second, the framework of structure theorems, as elucidated by the ergodic setting, unifies both the pure and impure proofs of Szemerédi's theorem. Though the association of explanation and unification has been pervasive in the philosophy of science (and more recently mathematics) literature,[121] not enough attention has been paid to what unification means in practice. One of the best discussions of this issue in the sciences is Morrison's book [48]. I believe that various distinctions made there can be carried over to the mathematical case and that these distinctions help to further the debate concerning unification and mathematical explanation. After suitably adapting some of these distinctions for the mathematical context, I conclude that, although unification (suitably understood) may not be associated with explanation in the sciences (as Morrison argues), the same cannot be said for mathematics. In particular, I extract

---

[118] See once more https://terrytao.wordpress.com/2012/03/23/some-ingredients-in-szemeredis-proof-of-szemeredis-theorem/.

[119] A generalized arithmetic progression is an arithmetic progression whose terms are themselves arithmetic progressions. Of course, after taking progressions of progressions (rank two), we can take progressions of progressions of progressions (rank 3), etc. The rank required for the combinatorial proof of Szemerédi's theorem is very large indeed (rank $2^k + 1$ where k indicates the length of the rank one progression we wish to find in the original set A stated in the theorem).

[120] Again, see my outline at the end of Section 3.1.

[121] See, for instance Friedman's classic article [21]; Kitcher's extensive work on the subject in [37–40]; a more recent discussion by Hafner and Mancosu [31]; and finally the chapters of [41] and [45] concerned with mathematical explanation.

a particular variety of unification from our case study, which I call *global synthetic unification*, and show how the comparative potential generated from this kind of unification serves to produce explanatory proofs.

To summarize: in this paper I have tried to motivate the investigation of an association between impure techniques of proof and explanation. I believe that these are more systematically connected in contemporary mathematics than has been acknowledged in the philosophical literature. I have also tried to defend this suggestion against the ancient and influential idea that purity and explanation are associated (or, even more strongly, necessarily connected). This has been done by isolating a central and important case in which radically impure techniques are brought to bear in order to provide an explanatory proof. I accounted for this phenomenon (and disarmed reasons for associating purity and explanation) by excavating a new notion of mathematical content, which I have called structural content. Finally, I sketched a proposal that, once this particular variety of content is in hand, one can see that impure resources lead to explanatory proofs via suitably understood varieties of simplification and unification. The full details of this final proposal must, however, wait for another day.

§6. Appendix: Excerpts from Szemerédi's proof.
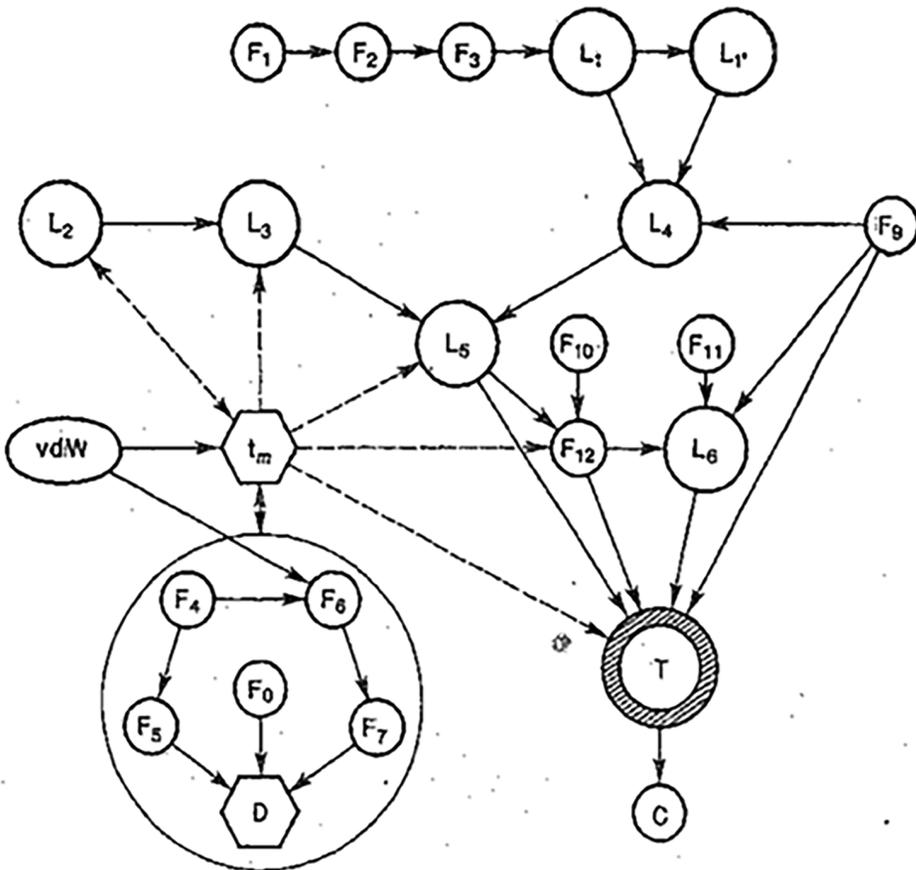
The diagram reprensents an approximate flow chart for the accompanying proof of Szemerédi's theorem. The various symbols have the following meanings: $F_k \equiv$ Fact k, $L_k \equiv$ Lemma k, $T \equiv$ Theorem, $C \equiv$ Corollary, $D \equiv$ Definitions of B, S, P, α, β, etc., $t_m \equiv$ Definition of t_m, vdW $\equiv$ van der Waerden's theorem, $F_0 \equiv$ "If $f : \mathbb{R}^+ \to \mathbb{R}^+$ is subadditive then $\lim_{n \to \infty} \frac{f(n)}{n}$ exists".

LEMMA 6.1 (Szemerédi Regularity Lemma; Lemma 1 in [59]). *Let A and B be disjoint sets. For $X \subseteq A$, $Y \subseteq B$ let I be a fixed subset of $[A, B] := (\{x, y\} : x \in X, y \in Y)$, $k(u) = \{v \in A \cup B : \{u, v\} \in I\}$, and $\beta(X, Y) = k(X, Y)|X|^{-1}|Y|^{-1}$. For all $\varepsilon_1, \varepsilon_2$, δ, ρ, σ, there exist m_0, n_0, M, N, such that for all I with $|A| = m > M$, $|B| = n > N$, there exist disjoint $C_i \subseteq A$, $i < m_0$, and, for each $i < m_0$, disjoint $C_{i,j} \subseteq B$, $j < n_0$, such that*:

1. *$|A - \cup_{i<m_0} C_i| < \rho m$, $|B - \cup_{j<n_0} C_{i,j}| < \sigma n$ for any $i < m_0$.*
2. *For all $i < m_0$, $j < n_0$, $S \subseteq C_i$, $T \subseteq C_{i,j}$, with $|S| > \varepsilon_1|C_i|$, $|T| > \varepsilon_2|C_{i,j}|$, we have*

$$\beta(S, T) \geq \beta(C_i, C_{i,j}) - \delta.$$

3. *For all $i < m_0$, $j < n_0$ and $x \in C_i$,*

$$|k(x) \cap C_{i,j}| \leq (\beta(C_i, C_{i,j}) + \delta)|C_{i,j}|.$$

BIBLIOGRAPHY

[1] Ackermann, W. (1937). Die Widerspruchsfreiheit der allgemeinen Mengenlehre. *Mathematische Annalen*, **114**, 305–315.

[2] Arana, A. (2015). On the depth of Szemerédi's theorem. *Philosophia Mathematica*, **23**(2), 163–176.

[3] ———. (2017). On the alleged simplicity of pure proof. In Kossak, R., and Ording, P., editors. *Simplicity: Ideals of Practice in Mathematics and the Arts*. Berlin: Springer, pp. 207–229.

[4] ———. (2019). Elementarity and purity. In Alvarez, C., and Arana, A., editors. *Analytic Philosophy and the Foundations of Mathematics*. London: Palgrave-Macmillan.

[5] Arana, A., & Mancosu, P. (2012). On the relationship between plane and solid geometry. *The Review of Symbolic Logic*, **5**(2), 294–353.

[6] Avigad, J. (2003). Number theory and elementary arithmetic. *Philosophia Mathematica*, **11**(3), 257–284.

[7] ———. (2009). The metamathematics of ergodic theory. *Annals of Pure and Applied Logic*, **157**, 64–76.

[8] Avigad, J., & Towsner, H. (2010). Metastability in the Furstenberg-Zimmer tower. *Fundamenta Mathematicae*, **210**, 243–268.

[9] Barnes, J. (1993). *Aristotle: Posterior Analytics*, second edition. Clarendon Aristotle Series. Oxford: Oxford University Press.

[10] Betti, A. (2010). Explanation in metaphysics and Bolzano's theory of ground and consequence. *Logique et Analyse*, **56**(211), 281–316.

[11] Bourbaki, N. (1950). The architecture of mathematics. *The American Mathematical Monthly*, **57**(4), 221–232.

[12] Burnyeat, M. (2012). Aristotle on understanding knowledge. In *Explorations in Ancient and Modern Philosophy*. Cambridge: Cambridge University Press, pp. 115–144.

[13] Dawson, J. W. (2006). Why do mathematicians re-prove theorems? *Philosophia Mathematica*, **3**(14), 269–286.

[14] de la Vallée Poussin, C. (1896). Recherches analytiques Sur la théorie des nombres premiers. *Annales de la Societe Scientifique de Bruxelles*, **20**, 183–256.

[15] Detlefsen, M. (2008). Purity as an ideal of proof. In Mancosu, P., editor. *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press, pp. 179–198.

[16] Detlefsen, M., & Arana, A. (2011). Purity of methods. *Philosophers' Imprint*, **11**(2), 1–20.

[17] Einsiedler, M., & Ward, T. (2011). *Ergodic Theory with a View towards Number Theory*. Graduate Texts in Mathematics, Vol. 259. Berlin: Springer-Verlag.

[18] Feferman, S. (1964). Systems of Predicative Analysis, I. *The Journal of Symbolic Logic*, **29**, 1–30.

[19] Field, H. (1980). *Science without Numbers*. Princeton: Princeton University Press.

[20] ———. (1989). *Realism, Mathematics and Modality*. Oxford: Basil Blackwell.

[21] Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, **71**(1), 1–19.

[22] Furstenberg, H. (1977). Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *Journal d' Analyse Mathématique*, **34**, 204–256.

[23] ———. (1981). Poincaré recurrence and number theory. *Bulletin of the American Mathematical Society*, **5**(3), 211–234.

[24] Furstenberg, H., Katznelson, Y., & Ornstein, D. (1982). The ergodic theoretical proof of Szemerédi's theorem. *Bulletin of the American Mathematical Society*, **7**(3).

[25] Furstenberg, H., & Weiss, B. (1978). Topological dynamics and combinatorial number theory. *Journal d' Analyse Mathématique*, **34**, 61–85.

[26] Gowers, T. (2001). A new proof of Szemerédi's theorem. *Geometric and Functional Analysis*, **11**, 465–588.

[27] ———. (2006). Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, **15**(1-2), 143–184.

[28] Green, B., & Tao, T. (2007). Szemerédi's theorem. *Scholarpedia*, **2**(7), 3446.

[29] ———. (2008). The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, **167**, 481–547.

[30] Hadamard, J. (1896). Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Société Mathématique de France*, **24**, 199–220.

[31] Hafner, J., & Mancosu, P. (2008). Beyond unification. In Mancosu, P., editor. *The Philosophy of Mathematical Practice*. New York: Oxford University Press, pp. 151–178.

[32] Hallett, M. (2008). Reflections on the purity of method in Hilbert's Grundlagen der Geometrie. In Mancosu, P., editor. *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press, pp. 198–255.

[33] Ingham, A.E. (2008). Review 10, 595c (mr0029411). *Bulletin of the American Mathematical Society*, **45**(4), 651–654.

[34] Isaacson, D. (1996). Arithmetical truth and hidden higher-order concepts. In Hart, W., editor. *The Philosophy of Mathematics*. New York: Oxford University Press, pp. 203–224.

[35] Kaye, R., & Wong, T. L. (2007). On interpretations of arithmetic and set theory. *Notre Dame Journal of Formal Logic*, **48**(4), 497–510.

[36] Kitcher, P. (1975). Bolzano's ideal of algebraic analysis. *Studies in History and Philosophy of Science*, **6**, 229–269.

[37] ———. (1976). Explanation, conjunction, and unification. *Journal of Philosophy*, **73**(8), 207–212.

[38] ———. (1981). Explanatory unification. *Philosophy of Science*, **48**, 507–531.

[39] ———. (1984). *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.

[40] ———. (1989). Explanatory unification and the causal structure of the world (sections 1-4.5). In Kitcher, P., and Salmon, W., editors, *Scientific Explanation*. Minnesota Studies in the Philosophy of Science, Vol. XIII. Minneapolis: University of Minnesota Press, pp. 410–437.

[41] Lange, M. (2017). *Because Without Cause*. Oxford: Oxford University Press.

[42] Mancosu, P. (1999). Bolzano and Cournot on mathematical explanation. *Revue d'histoire des sciences*, **52**(3/4), 429–455.

[43] ———. (2001). Mathematical explanation: Problems and prospects. *Topoi*, **20**, 97–117.

[44] ———. (2008a). Mathematical explanation: Why it matters. In Mancosu, P., editor. *The Philosophy of Mathematical Practice*. New York: Oxford University Press, pp. 134–150.

[45] ———. editor (2008b). *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press.

[46] Mancosu, P. (2018). Explanation in mathematics. Stanford Internet Encyclopedia of Philosophy. Available from: https://plato.stanford.edu/entries/mathematics-explanation.

[47] McLarty, C. (2010). What does it take to prove Fermat's last theorem? Grothendieck and the logic of number theory. *The Bulletin of Symbolic Logic*, **16**(3), 359–377.

[48] Morrison, M. (2000). *Unifying Scientific Theories*. Cambridge: Cambridge University Press.

[49] Nagle, B., Rödl, V., & Schacht, M. (2006). The counting lemma for regular k-uniform hypergraphs. *Random Structures and Algorithms*, **28**(22), 113–179.

[50] Ostwald, W. (1905). *Klassiker der Exacten Wissenschaften*, Vol. 153. Leipzig: Wilhelm, Engelmann.

[51] Rödl, V., & Schacht, M. (2007). Regular partitions on hypergraphs: Counting lemmas. *Combinatorics, Probability and Computing*, **16**(6), 887–901.

[52] Rödl, V., & Skokan, J. (2004). Regularity lemma for uniform hypergraphs. *Random Structures and Algorithms*, **25**(1), 1–42.

[53] ———. (2006). Applications of the regularity lemma for uniform hypergraphs. *Random Structures and Algorithms*, **28**(2), 180–194.

[54] Ross, W., & Minio-Paluello, L. (1964). *Aristotelis Analytica Priora et Posteriora*. Oxford: Oxford University Press.

[55] Rota, G.-C. (1997). *Indiscrete Thoughts*. Boston: Birkhäuser.

[56] Russ, S. (1980). A translation of Bolzano's paper on the intermediate value theorem. *Historia Mathematica*, **7**, 156–185.

[57] Simpson, S. (1999). *Subsystems of Second Order Mathematics*. Berlin: Springer.

[58] Steinkrueger, P. (2018). Aristotle on kind-crossing. *Oxford Studies in Ancient Philosophy*, **54**, 107–158.

[59] Szemerédi, E. (1975). On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, **XXVII**, 199–245.

[60] Tao, T. (2006a). The dichotomy between structure and randomness, arithmetic progressions, and the primes. *Proceedings of the International Congress of Mathematicians*, **I**, 581–608.

[61] ———. (2007). The ergodic and combinatorial approaches to Szemerédi's theorem. In Granville, A., Nathanson, M., and Solymosi, J., editors. *Additive Combinatorics*. Providence, RI: American Mathematical Society, pp. 145–193.

[62] ———. (2008). *Structure and Randomness: Pages from Year One of a Mathematical Blog*. Providence, RI: American Mathematical Society.

[63] Towsner, H. (2008). Some Results in Logic and Ergodic Theory. Ph.D. Thesis, Carnegie Mellon University.

[64] van der Waerden, B. (1928). Beweis einer Baudetschen Vermutung. *Nieuw Archief voor Wiskunde*, **15**, 212–216.

[65] ———. (1998). Wie der Beweis der Vermutung von Baudet gefunden wurde. *Elemente der Mathematik*, **53**, 139–148.

[66] Wittgenstein, L. (1967). *Remarks on the Foundations of Mathematics*. Cambridge, MA: M.I.T. Press.

[67] Zhao, Y. Szemerédi's Theorem via Ergodic Theory. Completed by the author in partial fulfillment of the requirements of the Part III Tripos in Mathematics at Cambridge University.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CA, USA
E-mail: patrick.j.ryan@berkeley.edu